Contributions au calcul dans les algèbres de Lie libres et à la déformation des groupes triangulaires en géométrie hyperbolique complexe

Pierre-Vincent Koseleff

UNIVERSITÉ PARIS 6

Habilitation à diriger des recherches

Spécialité : Mathématiques

Pierre-Vincent Koseleff

Titre :

Contributions au calcul dans les séries formelles de Lie et à la déformation de groupes triangulaires en géométrie hyperbolique complexe.

soutenue le 19 décembre 2003

devant le jury

Marc GIUSTI,
John PARKER,
Pierre PANSU, rapporteurs,

Daniel BENNEQUIN,
Gérard DUCHAMP,
Elisha FALBEL,
Jean-Jacques RISLER, examinateurs.

Table des Matières

Introduction **1**
Séries de Lie . . . 1
Parallélogrammes . . . 2
Déformation de groupes triangulaires dans $\mathbf{PU}(2,1)$ . . . 2
Problème d'Ising . . . 3

Présentation des résultats obtenus **5**
Problème général de l'approximation . . . 5
Intégrateurs symplectiques . . . 5
Planifications de trajectoires . . . 6
Parallélogrammes . . . 7
Séries de Lie . . . 7
Problème général de l'approximation . . . 8
Perspectives . . . 9
Groupes triangulaires de $\mathbf{PU}(2,1)$ . . . 10
Construction géométriques dans le modèle de Heisenberg . . . 11
Perspectives . . . 12

Liste des Travaux **13**

Bibliographie **15**

Annexes **19**
Exhaustive Search of Symplectic Integrators Using Computer Algebra . . . 21
Relations among Lie Series Transformations and Isomorphisms between free Lie Algebras . . . 43
Elementary Approximation of Exponential of Lie Polynomials . . . 57
The Number of Sides of a Parallelogram . . . 73
Rigidity and Flexibility of triangle groups in complex hyperbolic geometry . . . 85
A circle of Modular Groups in $\mathbf{PU}(2,1)$ . . . 107
Solving the triangular Ising ferromagnet by simple mean field . . . 121

INTRODUCTION

Ce mémoire aborde plusieurs domaines auxquels je me suis intéressés depuis quelques années : le calcul de Lie et en particulier les séries de Lie et leurs applications en théorie du contrôle (avec F. JEAN), en mécanique hamiltonienne et dans l'étude de relations dans des groupes ; l'étude des déformations de groupes triangulaires discrets dans l'espace $\mathbf{PU}(2,1)$ des automorphismes de la boule unité complexe de dimension 2 (avec E. FALBEL).
J'ai choisi dans ce mémoire, de présenter les résultats de quelques articles significatifs ainsi qu'un travail en collaboration avec Serge GALAM sur l'étude d'un modèle particulier du problème d'Ising triangulaire antiferromagnétique.

⋆⋆⋆

La caractéristique commune de ces travaux est le fait que les résultats sont donnés de façon explicite et constructive, utilisant les outils de calcul formel. À mon sens, c'est le point de vue algébrique, est plus particulièrement effectif, qui a permis d'obtenir les différents résultats que j'évoque succinctement ici et plus en détails dans la partie suivante.
Ainsi, dans la quasi totalité des références que je propose dans ce mémoire, toutes les solutions sont explicites et effectives, dans le sens que nous définissons un algorithme pour les calculer.

Séries de Lie

À l'origine, dans ma thèse ([13]), je m'étais intéressé aux méthodes de Lie en mécanique hamiltonienne. Elles jouent un rôle très important dans des domaines aussi variés que la mécanique céleste, l'optique géométrique, la physique des plasmas, la théorie des accélérateurs de particules, le transport de neutrons ou l'électricité. Leur essor remonte à la parution de deux articles, un article du mécanicien céleste André Deprit [110] en 1969 et celui de deux physiciens Alex Dragt et John Finn [111] en 1976.
Les mécaniciens célestes ont l'habitude de considérer les transformations canoniques comme des transformations de Lie, c'est-à-dire comme l'application, au temps 1, du flot d'un Hamiltonien non-autonome. Pour d'autres, on représente les transformations canoniques comme des composées d'applications, au temps 1, de flots hamiltoniens autonomes.
En fait, ces deux formalismes ne sont que deux représentations d'éléments du même groupe. L'étude (dans ma thèse) des liens entre ces transformations et l'exponentielle m'a conduit à développer la possibilité d'exprimer des identités entre les automorphismes des séries de Lie, dans une algèbre de Lie libre, tant d'un point de vue théorique que d'un point de vue effectif.
Les principaux résultats obtenus dans ma thèse étaient :

- Équivalence des transformations de Deprit, de Dragt-Finn et des exponentielles de dérivations intérieures
- Méthodes effectives de calcul des relations entre ces transformations et de leur composition (en particulier, obtention des formules de Campbell-Haussdorf à des ordres élevés).

– Obtention d'intégrateurs symplectiques : ce sont des schémas d'intégration numérique de systèmes hamiltoniens qui ont la particularité d'être sans dérive, c'est-à-dire, de conserver des intégrales premières proches de celles du système. Ces méthodes sont utilisées pour intégrer à très long terme des systèmes de mécanique céleste.

Développements postérieurs

J'ai poursuivi après ma thèse dans ce sujet et présenté plusieurs résultats :

– J'ai proposé dans *Exhaustive Search of Symplectic Integrators Using Computer Algebra* ([5]), une liste exhaustive d'intégrateurs symplectiques optimaux pour les petits ordres ainsi que des méthodes pour l'obtention à des ordres élevés. Outre l'intérêt de l'obtention de méthodes explicites d'intégration à très long terme de systèmes de mécanique céleste, l'obtention de ses résultats a mis en lumière des isomorphismes particuliers entre diverses composantes de l'algèbre de Lie libre.
– Utilisant des automorphismes des séries de Lie, j'ai mis en évidence (*Relations among Lie Series Transformations and Isomorphisms between free Lie Algebras,* [7]) certaines graduations de l'algèbre de Lie libre et certains isomorphismes qui m'ont conduit à démontrer que les composantes homogènes de la série de Hausdorff engendraient librement une algèbre de Lie libre (extension d'un résultat de Sirsov et Witt ([141])).
– Avec Frédéric JEAN (Ensta), considérant à présent les transformations de Lie, du point de vue de la théorie du contrôle, nous avons donné *Elementary Approximation of Exponential of Lie Polynomials* ([6]) une méthode explicite d'approximation à tout ordre de l'exponentielle d'un polynôme de Lie par un produit de facteurs élémentaires (généralisation de la formule de Zassenhauss).

Parallélogrammes

Comme je vais l'expliquer dans la partie suivante, les résultats évoqués ci-dessus ont comme caractéristique commune d'être des problèmes d'approximation dans un groupe de Lie, d'un élément par des éléments d'un sous-groupe particulier.

Ces méthodes explicites de calcul d'identités dans le groupe des transformations de Lie m'ont conduit à considérer, avec Elisha FALBEL (Paris 6), la question de la recherche d'identités de longueur minimale. Nos résultats concernent plus particulièrement les identités de longueur minimale (parallélogrammes) dans le groupe nilpotent libre. L'idée était de construire des trajectoires fermées comme successions d'orbites de champs de vecteurs. Dans (*The Number of Sides of a Parallelogram* [8]), nous bornons les longueurs des parallélogrammes et de nombreux exemples de groupes sont évoqués, soulignant ainsi la grande variété des cas possibles.

Déformation de groupes triangulaires dans $\mathbf{PU}(2,1)$

Une famille de groupes particulièrement intéressante est celle des groupes triangulaires. Je me suis intéressé (avec Elisha FALBEL) au problème de l'existence de la déformation de groupes

triangulaires discrets (engendré par 3 réflexions) dans le groupe des automorphismes de la boule complexe de dimension 2 : $\mathbf{PU}(2,1)$.
Nous nous sommes particulièrement intéressés au groupe $(2,3,\infty)$ dont $\mathbf{SL}(2,\mathbb{Z})$ peut être vu comme un sous-groupe d'indice 2.
Notre travail, outre qu'il donne des résultats de flexibilité (existence de familles de déformation) là où existaient des résultats de rigidité, est fondé sur l'utilisation des **C**-sphères (mis en évidence par Falbel et Zocca [115]). Ce sont des surfaces qui délimitent des domaines fondamentaux, lesquels permettent de conclure sur le caractère discret du groupe étudié. Une part importante du travail a été l'étude d'invariants algébriques intervenant dans la construction de tels objets. Ce sont ces constructions ([9, 10, 12]), qui ont nécessité l'utilisation du calcul formel, qui ont permis de mettre à jour des familles de déformations.

Problème d'Ising

Avec Serge GALAM (CNRS, Paris 6), nous avons mené ([11]) une étude algébrique et donc basée sur des méthodes de calcul formel pour le modèle d'Ising antiferromagnétique.
Le modèle d'Ising permet d'étudier les transitions de phase des systèmes de la physique de la matière condensée. Ici nous étudions une nouvelle théorie de champ moyen, proposée par S. Galam ([118, 119]) qui préserve la symétrie hamiltonienne initiale. Ce modèle particulièrement simple est appliqué ici pour résoudre le modèle antiferromagnétique triangulaire d'Ising. Ce modèle a ceci d'intéressant que les résultats que nous obtenons ne sont pas numériques. L'existence des états d'équilibre et leur détermination a pu se démontrer de façon exacte (algébrique). Là où tous les modèles de champ moyen indiquaient une transition de phase à température non nulle, nous avons montré qu'il n'en était rien avec ce modèle particulier, conformément à la théorie générale ([151]).
Ce modèle simple de champ moyen ouvre une nouvelle manière d'aborder les systèmes aléatoires.

⋆⋆⋆

PRÉSENTATION DES RÉSULTATS OBTENUS

Outre le travail en commun avec S. Galam (*Solving the triangular Ising ferromagnet by simple mean field*), qui concerne l'utilisation de méthodes algébriques simples et classiques dans l'étude d'un modèle mathématique particulier du problème d'Ising antiferromagnétique, on peut séparer les thèmes de recherche dans les travaux que je présente en deux grandes parties : les **séries de Lie** et la **géométrie hyperbolique complexe**. Ces deux thèmes sont liés par le premier travail que j'ai effectué avec E. Falbel : *The Number of Sides of a Parallelogram*. Il s'agissait au départ de construire des trajectoires fermées comme successions d'orbites de champs de vecteurs particuliers. Il s'est vite avéré qu'un problème plus général était celui de la recherche d'identités de longueur minimale dans un groupe (de Lie) donné ou même plus généralement le problème de l'approximation d'un élément d'un groupe par des éléments d'un sous-groupe. Le cadre dans lequel nous nous sommes placés dans notre étude est celui du groupe nilpotent libre, ce qui, en utilisant diverses graduations, peut conduire à des résultats plus généraux d'approximation dans le cadre des séries formelles de Lie.

Problème général de l'approximation

Soit x, y des éléments d'une algèbre de Lie L. Dans le groupe de Lie $\exp(L)$, nous savons que

$$\exp(x+y) \simeq \exp(x)\exp(y),\, 1 \simeq \exp(x)\exp(y)\exp(-x)\exp(-y)$$

dans un sens qu'il convient de préciser. Ce sont là les premiers exemples d'approximations. Dans le premier cas nous approximons un élément (l'exponentielle d'une somme) par un produit d'éléments d'un sous groupe et dans le second cas nous recherchons une approximation de 1. Encore ne s'agit-il ici que d'approximation du premier ordre.

Intégrateurs symplectiques

Les intégrateurs symplectiques sont des schémas d'intégration numérique à long terme, de systèmes hamiltoniens. Le problème est le suivant : étant donné un hamiltonien $H = H_1 + H_2$, calculer une solution approchée de l'équation

$$z(0) = z_0, \dot{z}(t) = [H, z].$$

Formellement, la solution est donnée par $z(t) = \exp(t[H,\cdot])z_0$, lorsque H ne dépend pas du temps. Il est fréquent que $\exp(t[H,\cdot])$ ne se calcule pas facilement alors que $\exp(t[H_1,\cdot])$ et $\exp(t[H_2,\cdot])$ le peuvent. L'idée est d'essayer d'approximer le flot $\exp(\tau[H,\cdot])$ par une composition des flots hamiltoniens $\exp(\tau_1[H_1,\cdot])$ de H_1 et $\exp(\tau_2[H_2,\cdot])$ de H_2. De tels intégrateurs peuvent se construire en considérant des identités universelles dans les algèbres de Lie libres :

$$\exp(\tau[H,\cdot]) \simeq \exp(\tau_1[H_{i_1},\cdot])\cdots\exp(\tau_k[H_{i_k},\cdot])$$

de telle façon que l'erreur ne soit pas importante.
L'avantage est double : tout d'abord, les méthodes classiques d'intégrations numériques ne sont pas sans dérive, c'est-à-dire, l'intégrale première $H(z)$, non seulement ne sera pas constante mais de plus va tendre vers l'infini ; ici, l'approximation ainsi obtenue du flot hamiltonien apparaît comme une combinaison de flots hamiltoniens et est enoce un flot hamiltonien.
Cette approche a été introduite en 1988. L'utilisation du formalisme des séries de Lie pour résoudre cette question est apparue au début des années 1990 [117, 153, 154, 146, 147, 148, 134]. Dans ces travaux, des méthodes de construction pour des petits ordres ou pour des ordres élevés ne prouvaient pas leur exhaustivité ni même leur minimalité. Dans ma thèse ([13]) et dans deux articles ([2, 5]) je donne une liste exhaustive de tels intégrateurs pour des petits ordres et propose des constructions pour des ordres élevés.
Dans le cas général, il s'agit de se placer dans un certain groupe de transformations. Le cadre le plus général est celui des automorphismes des séries de Lie dans lequel j'avais déjà étudié la représentation des compositions.
Dans le cas particulier de l'intégration de systèmes de mécanique céleste, mes résultats, qui utilisaient le fait que le Hamiltonien du système pouvait se ré-écrire, dans les variables héliocentriques canoniques, comme la somme de problèmes à deux corps et une perturbation d'interaction, nous avions de plus une relation de la forme

$$[[[T,V],V],V] = 0,$$

due au fait que l'énergie cinétique du système est une forme quadratique. Je propose alors dans *Exhaustive Search of Symplectic Integrators Using Computer Algebra*, une construction en me plaçant dans le cadre d'un quotient d'algèbre de Lie libre.
Ces méthodes sont utilisées actuellement en mécanique céleste (Laskar et Robutel [130]), ou plus généralement en intégration de systèmes "symplectiques" (McLachlan [135]).

Planifications de trajectoires

En théorie du contrôle, pour un système $\dot{x} = \sum_{i=1}^{m} u_i(t)X_i(x)$, le problème classique de la planification de trajectoires est ([126, 129]) : *étant donnés deux états p et q, déterminer une trajectoire réalisable (i.e. des contrôles $u_1(t),\dots,u_m(t)$) tels que $x(0) = p$ et $x(1)$ soit arbitrairement proche de q.*
Nous obtenons une trajectoire comme une concaténation de trajectoires, chacune étant l'orbite pendant le temps λ_{i_k} sous l'action du champ X_{i_k}.
Soit q un état donné comme $\exp(X)p$, dans lequel X appartient à l'algèbre de Lie $\mathcal{L}(X_1,\dots,X_m)$, engendrée par les champs de vecteurs X_i. Nous recherchons des trajectoires, plus simples, comme composition des flots des X_i.
Le point final de la trajectoire est donné alors par $\exp(\lambda_1 X_{i_1})\cdots\exp(\lambda_s X_{i_s})p$, c'est-à-dire, un produit de facteurs élémentaires appliqué à l'état p. Ici encore, il s'agit d'écrire

$$\exp(tX) \simeq \exp(\lambda_1 X_{i_1})\cdots\exp(\lambda_s X_{i_s}) = \exp(tX + o(t^k)).$$

Dans (*Elementary Approximation of Exponentials of Lie Polynomials*, [6]), nous donnons une méthode effective qui détermine de telles trajectoires à tout ordre. Ce ne sont pas des identités minimales. Elles utilisent tout d'abord la décomposition en forme factorisée (prop. 1) puis pour chaque terme homogène une méthode sans calculs dans l'algèbre de Lie.

Parallélogrammes

Le cas particulier où les $\lambda_i \in \mathbb{Z}$ (resp. $\{-1,1\}$) et $X = 0$ a été étudié dans [8]. Nous l'avons appelé parallélogramme et il est lié au problème suivant : *soit a,b des éléments d'un groupe G. Un parallélogramme est une relation minimale de la forme*

$$1 = a^{n_1} b^{m_1} \cdots a^{n_k} b^{m_k}$$

Ici nous cherchons à minimiser la longueur $\sum_{i=1}^{k} |n_i| + |m_i|$.

Dans le cas du groupe nilpotent libre d'ordre m, nous montrons que le problème est identique à celui de la recherche d'un élément

$$\exp(n_1 x)\exp(m_1 y)\cdots\exp(n_k x)\exp(m_k y) - 1$$

d'ordre m dans l'algèbre associative libre $L(x,y)$, ou de la série rationnelle

$$(1+x)^{n_1}(1+y)^{m_1}\cdots(1+x)^{n_k}(1+y)^{m_k} - 1$$

d'ordre m.

⋆⋆⋆

Pour ces trois problèmes d'approximation, nous nous sommes placés dans la cadre des algèbres de Lie libres ou du groupe nilpotent libre. Les résultats obtenus ont utilisés le formalise des automorphismes des séries de Lie et des relations explicites que nous pouvions obtenir entre diverses transformations. Les résultats explicites ont été obtenus grâce à des outils de calculs formel que j'ai continué de développer peu après ma thèse et par des techniques de résolution de systèmes polynomiaux.

Séries de Lie

X étant un alphabet (éventuellement pondéré), $L(X), F(X)$ et $\mathcal{A}(X)$ désignent l'algèbre de Lie libre, le groupe libre et l'algèbre associative libre sur X. Ils sont classiquement gradués par la longueur, le poids (longueur pondérée) et le multi-degré.
On définit les séries formelles de Lie $\widehat{L}(X)$ et les séries non commutatives $\widehat{\mathcal{A}}(X)$ par complétion.

Nous utilisons aussi $\widehat{L}_{\geq p}(X) = \prod_{n \geq p} L_n(X)$, et $\widehat{\mathcal{A}}_{\geq p}(X) = \prod_{n\geq p} \mathcal{A}_n(X)$ où $L_n(X)$ (resp $\mathcal{A}_n(X)$) est le sous-module des éléments de longueur n.

L'ensemble $\Gamma(X) = 1 + \widehat{\mathcal{A}}_{\geq 1}(X)$ est appelé le *groupe de Magnus*.

Exponentielle.— Ayant définit l'exponentielle $\exp : \widehat{\mathcal{A}}_{\geq 1}(X) \to \Gamma(X)$ nous utilisons de théorème de Campbell-Hausdorff [105, Ch. II, §5]) :

Théorème 1 (Campbell-Hausdorff).— *Pour $x, y \in \widehat{L}_{\geq 1}(X)$, la série de Hausdorff $H(x,y) = \log[\exp(x)\exp(y)]$ appartient à $\widehat{L}_{\geq 1}(X)$.*

Une conséquence directe est une variante de la formule de Zassenhauss [141] :

Proposition 1 (Développement en produit).— *Pour $k \in \widehat{L}_{\geq 1}(X)$, il existe un unique $g \in \widehat{L}_{\geq 1}(X)$ tel que*

$$\exp(\sum_{n\geq 1} k_n) = \cdots \exp(g_n) \cdots \exp(g_1).$$

De plus, on peut calculer explicitement g en fonction de k (voir par exemple ma thèse [13]).

Problème général de l'approximation

Le problème général de l'approximation est le suivant :

Problème 1 (Approximation).— *Soit $X = \{X_1, \ldots, X_k\}$, un alphabet. Soit $P \in L(X)$. Déterminer $\lambda_1, \ldots, \lambda_s$ dans un ensemble à préciser $(\mathbb{R}, \mathbb{C}, \mathbb{Z}, \{-1, 1\})$, tels que*

$$\exp(P) = \exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s}) \exp(R_{>n})$$

où les $X_{i_k} \in X$ et $R_{>n} \in \widehat{L}_{\geq n+1}(X)$.

Parallélogrammes [8]

Notons $F_{\geq 1}(X) = F(X)$ et posons $F_{\geq n}(X) = (F_{\geq 1}(X), F_{\geq n-1}(X))$ (ensemble des commutateurs), on obtient la *suite centrale descendante*.

Considérant les filtrations centrales de $F(X) \to \Gamma(X)$, $\mu : x \in X \mapsto (1+x)$ et $\mu' : x \in X \mapsto \exp(x)$, Magnus a prouvé le résultat ([105]) : $\mu^{-1}(1 + \widehat{\mathcal{A}}_{\geq n}(X)) = \mu'^{-1}(1 + \widehat{\mathcal{A}}_{\geq n}(X)) = F_{\geq n}(X)$
Cette propriété nous permet de considérer les polygones comme des approximants

$$\begin{array}{rcl} \exp(a_1 x)\exp(b_1 y) \cdots \exp(a_n x)\exp(b_n y) & \in & 1 + \widehat{\mathcal{A}}_{\geq m}(X), \\ (1+x)^{a_1}(1+y)^{b_1} \cdots (1+x)^{a_n}(1+y)^{b_n} & \in & 1 + \widehat{\mathcal{A}}_{\geq m}(X). \end{array}$$

ce que l'on peut à la fois rapprocher des séries de Lie et des séries rationnelles (voir [102]). Ici, les a_i et les b_i sont des entiers non nuls. Un parallélogramme d'ordre m sera un polygone d'ordre m de longueur minimum l_m et nous démontrons, utilisant un minorant du rang d'une série rationnelle : [8]

Théorème 2.— *On a $m \leq l_m \leq m^2$.*

Nous donnons une méthode explicite pour construire des parallélogrammes à tout ordre et étudions divers cas particuliers lorsque le groupe G n'est pas le groupe nilpotent libre.

Intégrateurs symplectiques [4, 5].
Le cas particulier où $P = X_1 + \cdots + X_k$ (en particulier $k = 2$) concerne la théorie des intégrateurs symplectiques

$$\exp(P) = \exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s}) \exp(R_{>n})$$

Ici les λ_i sont des nombres algébriques (réels ou complexes) et il est important de pouvoir estimer l'erreur $\exp(R_{>n})$. Les X_i sont des champs de vecteurs hamiltoniens $[H_i, \cdot]$. Nous écrivons

$$H = H_1 + \cdots + H_k,$$

de telle façon que l'intégration numérique ou exacte des flots des hamiltoniens H_i soit connue.

Théorie du contrôle [6].
Nous avons montré : pour tout $P \in L(X_1, \ldots, X_m)$, pour tout $m \in \mathbb{N}$, il existe $\lambda_1, \ldots, \lambda_s$ réels, tels que

$$\exp(P) = \exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s}) \exp(R_{>n}).$$

Les λ_i, sont donnés de façon explicite, sans aucun calcul, si P est un monôme de Lie. Sinon, nous utilisons la décomposition de la proposition 1 qui est effective.

⋆⋆⋆

Le problème général de l'approximation a toujours une solution. Une question intéressante est également : *quel est l'ensemble des solutions* $\lambda_1, \ldots, \lambda_s$ *de*

$$\exp(P) = \exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s}) \exp(R_{>n}).$$

J'ai montré, dans ma thèse, que les s-uplets $(\lambda_1, \ldots, \lambda_s)$ sont les zéros d'un certain idéal polynomial qu'il est possible de déterminer. Celui-ci dépend de la base qui a été fixée dans l'algèbre de Lie et de la façon de représenter les transformations.
En particulier, cherchant des relations entre diverses transformations qui sont des automorphismes de $\widehat{L}(X)$, et divers isomorphismes entre sous-modules de cette algèbre, j'ai pu montrer que les composantes homogènes de la série de Hausdorff engendraient librement $L(X)$.

⋆⋆⋆

Perspectives

Je me suis écarté depuis 1997 des applications telle que la planification de trajectoires ou la recherche d'intégrateurs symplectiques. Néanmoins, bien que les numériciens ou les utilisateurs de tels intégrateurs aient découvert que ceux obtenus avec des valeurs négatives aient de moins bonnes propriétés de stabilité, une piste n'a pas été explorée avec les intégrateurs utilisant des valeurs complexes. Il me semble que cette piste pourrait être étudiée, à condition de donner un sens à de tels méthodes.
Concernant l'étude des parallélogrammes, les quelques exemples que nous avons illustrés montrent à l'évidence que certaines propriétés restent à être découvertes. Peut-être la longueur des parallélogrammes nous renseignerait-elle sur la structure des groupes. C'est en tout cas une piste que je compte bien reprendre.

Groupes triangulaires de $\mathbf{PU}(2,1)$

Un problème de base en géométrie est celui de la déformation. Partant d'un groupe Γ abstrait finiment engendré, et d'un groupe de Lie G_1, on peut rechercher un plongement $\rho_0 : \Gamma \to G_1$ et se demander si celui-ci ne fait partie d'une famille ρ_t de plongements discrets.
Un cas particulièrement intéressant est le groupe triangulaire de type (p,q,∞), présenté par

$$\Gamma =< \iota_0, \iota_1, \iota_2 : {\iota_0}^2 = 1, \iota_1^2 = 1, \iota_2^2 = 1, (\iota_0 \circ \iota_1)^p = 1, (\iota_0 \circ \iota_2)^q = 1 > .$$

Prolongeant le travail initié par Falbel et Zocca [115], nous avons étudié des déformations du plongement de ce groupe triangulaire dans le groupe des isométries de la boule unité complexe de dimension 2. Ce sont les applications conformes du compactifié du groupe de Heisenberg **H**, lequel peut s'identifier à S^3, le bord du plan hyperbolique complexe.
Il était déjà connu, qu'un plongement dans d'autres groupes d'isométries d'espaces symétriques, tels le disque unité de dimension 1, ou l'espace hyperbolique réel de dimension 3, est rigide.
Dans le cas auquel nous nous intéressons, nous obtenons des déformations non triviales, pour des groupes de réflexions Γ triangulaires (p,q,∞). Celles-ci sont construites en exhibant des domaines fondamentaux comme "polyèdres" dont la frontière est une **C**-sphère, c'est-à-dire, est feuilleté par des **C**-cercles, utilisant un théorème de Poincaré modifié (Falbel, Zocca [115]).
Les techniques utilisées par Goldman [121] ou Parker et Goldman [123] ou Parker et Gusevskii [124] utilisaient des bisecteurs (hypersurfaces feuilletées par des **C**-cercles), qui offrent moins de liberté.
Citons quelques résultats significatfs

Théorème 3 (Goldman [122]).**—** *Si $\rho_0(\Gamma)$ laisse invariante une variété totalement géodésique complexe et si son domaine fondamental, restreint à cette variété, est compact, alors toute déformation proche de ce plongement est conjuguée à celui-ci.*

Dans *Rigidity and Flexibility of triangle groups in complex hyperbolic geometry*, nous montrons le résultat suivant :

Théorème 4 (Falbel & Koseleff [10]).— *Il existe un plongement discret et fidèle du groupe* $\Gamma(p,q,\infty)$*, qui fixe une variété totalement géodésique complexe, et admettant un voisinage (de dimension 4) de plongements discrets et fidèles.*

Dans *A circle of modular groups in* $\mathbf{PU}(2,1)$, nous montrons :

Théorème 5 (Falbel & Koseleff [12]).— *Il existe une famille à un paramètre* $\rho_t(\Gamma(2,3,\infty))$ *de plongements discrets, telle que* $\rho_0(\Gamma)$ *laisse invariante une variété totalement géodésique complxe et* $\rho_1(\Gamma)$ *fixe une variété totalement géodésique réelle. Les groupes* $\rho_t(\Gamma)$ *sont engendrés par des réflexions réelles.*

Construction géométriques dans le modèle de Heisenberg

Nous nous plaçons dans la formalisme du groupe de Heisenberg **H** (voir [122, 115, 9, 10]). C'est l'ensemble $\mathbb{C}\times\mathbb{R}$ que nous identifions au bord de la boule complexe, muni de l'addition $(z,t)+(z',t')=(z+z',t+t'+2\mathrm{Im}\, z\bar{z}')$. **H** s'identifie à la frontière du plan hyperbolique complexe $\mathbf{H}^2{}_C$. Dans $\mathbf{H}^2{}_C$, il y a deux familles de surfaces totalement géodésiques : ce sont les sous-variétés totalement réelles (isométriques de $\mathbf{H}^2{}_C\cap\mathbb{R}^2$) et les sous-variétés totalement complexes (isométriques de $H^2{}_C\cap\mathbb{C}$).
Dans le modèle de Heisenberg, nous considérons les intersections de ces plans géodésiques avec S^3 pour obtenir les **R**-cercles et les **C**-cercles.
Le groupe $\mathbf{PU}(2,1)$ dans lequel nous plongeons Γ contient des transformations holomorphes et antiholomorphes. Il y a deux types de réflexions dans ce modèle : les réflexions réelles qui sont toutes conjuguées à l'inversion (anti-involution) standard

$$I_0 : (z,t) \mapsto \left(\frac{-z}{|z|^2-it}, -\frac{t}{|z|^4+t^2}\right),$$

qui fixe point par point la courbe $\mathbf{R}_0 : r^2+it=-e^{-2i\theta}$, laquelle se projette sur le plan $\mathbb{C}$ en la lemniscate de Bernoulli. $\mathbf{R}_0$ est le **R**-*cercle standard*.
L'autre anti-involution particulière est $I_1 : (z,t)\mapsto(\bar{z},-t)$, laquelle fixe la droite $\mathbf{R}_1$ (l'axe réel).
Dans le modèle de Heisenberg, les **R**-cercles (intersection entre S^3 et une sous-variété totalement réelle) sont soit des droites (**R**-cercles infinis), transformés de $\mathbf{R}_1$ par des similitudes, soit des transformés de $\mathbf{R}_0$ par des similitudes du groupe de Heisenberg (**R**-cercles finis).
À chaque **R**-cercle, nous pouvons associer une unique inversion. Les objets laissés globalement invariants par ces inversions sont des **C**-cercles. Dans le modèle de Heisenberg, les **C**-cercles sont soit des droites verticales, soit des ellipses, intersection d'un cylindre vertical à base circulaire et d'un plan.

Pour démontrer que la configuration est un groupe discret, nous construisons explicitement un domaine fondamental dont la frontière sera une réunion de **C**-cercles (ellipses), utilisant le théorème de Poincaré [142, 137, 115]. Nous ramenons donc le problème à celui de la construction de familles à un paramètre de **C**-cercles, lesquelles doivent s'intersecter suivant des **C**-cercles invariants par deux des générateurs.
Nous avons étudié les configurations possibles et proposé plusieurs constructions qui s'appuyaient essentiellement sur les propriétés algébriques de ces ellipses.

⋆⋆⋆

Mon apport principal à ce travail a été de mettre en évidence des propriétés algébriques des **C**-sphères, permettant ainsi la construction explicite de domaines fondamentaux.
Bien que celà n'apparaissent pas explicitement dans nos travaux, les résultats ont été obtenus grâce à l'utilisation du calcul formel : soit pour montrer qu'une construction était impossible (en utilisant des techniques de géométrie algébrique réelle de dénombrements de zéros), soit pour donner des formules explicites. En particulier, j'ai développé un *paquetage* dans le logiciel MAPLE de factorisation des séries de Poisson (combinaisons linéaires de fonctions trigonométriques).

Perspectives

Sur la déformation des plongements des groupes triangulaires (p,q,∞), de nombreuses questions restent en suspens. Nous savons déja, que pour tous p et q, il existe des voisinages d'une configuration que nous pouvons déformer non trivialement. Par contre nous ne savons pas montrer (parce que nous ne savons pas construire de domaine fondamental) qu'au delà, certaines déformations ne sont plus discrètes. Il reste néanmoins de nombreuses possibilités d'étude, en étudiant des invariants algébriques des générateurs, qui s'appuierait sur des techniques de la théorie des invariants.

LISTE DES TRAVAUX

Articles

[1] KOSELEFF, P.-V., *Jeux de mots dans les algèbres de Lie libres : quelques bases et formules*, Theoret. Comput. Sci. **79, no. 1, (Part A)** (1991), 241–256

[2] KOSELEFF, P.-V., *Relations among Formal Lie Series and Construction of Symplectic Integrators*, AAECC'10 proceedings, Lect. Not. Comp. Sci. **673** (1993), 213–230

[3] KOSELEFF, P.-V., *Comparison between Deprit and Dragt-Finn perturbation methods*, Celestial Mechanics **58-1**, Kluwer (1994)

[4] KOSELEFF, P.-V., *About approximations of exponentials*, Lecture Notes in Computer Science **948**, Springer (1995), 323–333

[5] KOSELEFF, P.-V., *Exhaustive Search of Symplectic Integrators Using Computer Algebra*, Fields Institute Communications **10** (1996)

[6] JEAN, F., KOSELEFF, P.-V., *Elementary Approximation of Exponentials of Lie Polynomials*, Lect. Not. Comp. Sci. **1255**, Springer (1997), 210–230

[7] KOSELEFF, P.-V., *Relations among Formal Lie Series Transformations and isomorphisms between free Lie algebras*, Discrete Mathematics **180** (1998), 243–254

[8] FALBEL, E., KOSELEFF, P.-V., *The Number of Sides of a Parallelogram*, Discrete Mathematics and Theoretical Science, **3-2** (1999), 33–42

[9] FALBEL, E., KOSELEFF, P.-V., *Flexibility of the ideal triangle group in complex hyperbolic geometry*, Topology **39(6)**, Pergamon (2000), 1209–1223

[10] FALBEL, E., KOSELEFF, P.-V., *Rigidity and Flexibility of triangle groups in complex hyperbolic geometry*, Topology **41(4)**, Pergamon (2002), 767–786

[11] GALAM, S., KOSELEFF, P.-V., *Solving the triangular Ising ferromagnet by simple mean field*, The European Physical Journal B **28** (2002), 149–155

[12] FALBEL, E., KOSELEFF, P.-V., *A circle of modular groups in* $\mathbf{PU}(2,1)$, Mathematical Research Letters **9** (2002), 379–394

Thèse, livres

[13] KOSELEFF, P.-V., *Calcul Formel pour les méthodes de Lie en mécanique hamiltonienne*, Thèse de troisième cycle, École Polytechnique, 1993

[14] KOSELEFF, P.-V., *Formal Calculus for Lie Methods in Hamiltonian Mechanics* **LBID-2030 Rev UC 405**, Lawrence Berkeley Laboratory, 1994

[15] JACOB, G., KOSELEFF, P.-V. (ÉDS), *Special issue : 'Lie Computations'*, Discrete Mathematics and Theoretical Science, **1-1**, Springer, 1997

[16] KOSELEFF, P.-V., *Configurations Centrales de quatre corps dans le plan*, in *Modélisation mathématique, un autre regard*, Scopos **16**, Springer (2002), 115–124

BIBLIOGRAPHIE

[101] BELLAÏCHE A., RISLER, J.-J. (EDS), *Sub-Riemannian Geometry,* Progress in Mathematics **144**, Birkhäuser, 1996

[102] BERSTEL, J., REUTENAUER, C., *Rational series and their languages,* EATCS Monographs on Theoretical Computer Science, Springer, 1988

[103] BEARDON, A. F., *The Geometry of Discrete groups*, Springer-Verlag, 1983

[104] BURNS, D., SHNIDER, S., *Spherical Hypersurfaces in Complex Manifolds,* Invent. Math. **33** (1976), 223–246

[105] BOURBAKI, N., *Groupes et algèbres de Lie*, Éléments de Mathématiques, Hermann, Paris, 1972

[106] CARTAN, E., *Sur le groupe de la géométrie hypersphérique*, Comm. Math. Helv. **4** (1932), 158–171

[107] CARTIER, P., *Démonstration algébrique de la formule de Hausdorff*, Bull. S.M.F. **84** (1956), 241–249

[108] CARY, J.R., *Lie Transform Perturbation Theory for Hamiltonian Systems*, Physics Reports **79-2**, North-Holland Publishing Company (1981), 129–159

[109] S. CHEN, L. GREENBERG, *Contributions to Analysis*, a, Academic Press, New York, 1974

[110] DEPRIT, A., *Canonical transformations depending on a small parameter*, Cel. Mech. **1** (1969), 12–30

[111] DRAGT, A. J., FINN, J. M., *Lie Series and invariant functions for analytic symplectic maps*, J. Math. Physic **17** (1976), 2215–2227

[112] DRAGT, A. J., HEALY, L. M., *Lie Methods in Optics II,* Lec. Notes in Physics **352**, Springer Verlag, 1988

[113] FALBEL, E., GORODSKI, C., RUMIN, M., *Holonomy of sub-Riemannian manifolds*, International Journal of Mathematics **8, No. 3** (1997), 317–344

[114] FALBEL, E., PARKER, J., R., *The moduli space of the modular group in complex hyperbolic geometry*, Invent. Math. **152(1)** (2003), 57—88

[115] FALBEL, E., ZOCCA, V., *A Poincaré's polyhedron theorem for complex hyperbolic geometry*, J. Reine Angew. Math. **5126**, Walter de Gruyter (1999), 133–158

[116] FINN, J. M., *Lie Series : a Perspective, Local and Global Methods of nonlinear Dynamics*, Lec Notes in Physics **252**, Springer Verlag (1984), 63–86

[117] FOREST, E., RUTH, D., *Fourth-Order Symplectic Integration,* Physica D **43**, Elsevier Sc. Publ. B.V. (North-Holland) (1990), 105–117

[118] GALAM, S., *Self-consistency and symmetry in d dimensions*, Phys. Rev. B **54** (1996), 15991–15996

[119] GALAM, S., *From Galam-Mauger law to a powerful mean field scheme*, Journal of Applied Physics **87 (9 :3)** (2000), 7040–7042

[120] GOLDMAN, W., *Geometry and Topology Proceedings, University of Maryland 1983-1984*, Lecture Notes in Mathematics **1167**, Springer

[121] GOLDMAN, W., MILLSON, J.-J., *Local rigidity of discrete groupes acting on complex hyperbolic space*, Inventiones Mathematicæ **88** (1987), 495–520.

[122] GOLDMAN, W., *Complex Hyperbolic Geometry*, Oxford Mathematical Monographs, Oxford Science Publications, 1999

[123] GOLDMAN, W., PARKER, J., *Complex hyperbolic ideal triangle groups*, J. Reine Angew. Math. **425** (1992), 71–86

[124] GUSEVSKII, N., PARKER, J. R., *Representations of free Fuchsian groups in complex hyperbolic space*, Topology **39** (2000), 33–60

[125] HITCHIN, N. J., *Lie Groups and Teichmüller Space*, Topology **31(3)** (1992), 449–473

[126] JACOB, G., *Motion Planning by piecewise constant or polynomial inputs*, Proceedings of the IFAC Nonlinear Control Systems Design Symposium, Bordeaux, France (1992)

[127] A. KORÁNYI, H. M. REIMANN, *Quasiconformal mappings on the Heisenberg group*, Invent. Math. **80** (1985), 309–338

[128] LABUTE, J.-P., *Groups and Lie algebras : the Magnus theory*, in *mathematical legacy of Wilhelm Magnus : groups, geometry and special functions*, Contemp. Math. **169**, A.M.S. (1992)

[129] LAFFERRIERE, G., SUSSMANN H., *Motion Planning for controllable systems without drift*, Proceedings of the 1991 IEEE International Conference on Robotics and Automation, Sacramento, California (1991)

[130] LASKAR, J., ROBUTEL, PH., *High order symplectic integrators for perturbed Hamiltonian systems*, Celestial Mech. Dynam. Astronom **80(1)** (2001), 39–62

[131] LAUMOND, J.P., *Nonholonomic Motion Planning via Optimal Control*, Algorithmic Foundations of Robotics, A.K. Peters Pub. (1995)

[132] LYNDON, R. C. SHUPP P. E., *Combinatorial group theory*, Springer, 1977

[133] MCLACHLAN, R. I., *On the numerical integration of ordinary differential equations by symmetric composition methods*, SIAM J. Sci. Comp. **16(1)** (1995), 151–168

[134] MCLACHLAN, R. I., SCOVEL, C, *Open problems in symplectic integration*, in *Integration Algorithms and Classical Mechanics*, Fields Institute Communications **10**, AMS, J.E. Marsden, G.W. Patrick, and W.F. Shadwick, eds. (1996), 151–180

[135] MacLachlan, R. I., *Families of high-order composition methods*, Numerical Algorithms **31** (2002), 233–246

[136] Magnus *et al*, *Combinatorial Group Theory : Presentation of Groups in Terms of Generators and Relations*, J. Wiley & Sons, 1966

[137] Maskit, B., *Kleinian Groups*, Springer-Verlag, 1988

[138] Michel, J., *Bases des Algèbres de Lie Libres, Étude des coefficients de la formule de Campbell-Hausdorff*, Thèse, Orsay, 1974

[139] Perrin, D., *Factorization of free monoids*, in *Lothaire M., Combinatorics On Words, Chap. 5*, Addison-Wesley (1983)

[140] Petitot, M., *Algèbre non commutative en Scratchpad : application au problème de la réalisation minimale analytique*, Thèse, Université de Lille I, 1991

[141] Reutenauer, C., *Free Lie algebras*, Oxford Science Publications, 1993

[142] de Rham, G., *Sur les polygones générateurs de groupes fuchsiens*, Enseign. Math. **17** (1971), 49–61.

[143] Sanov, I. N., *A property of a representation of a free group*, Dokl. Akad. Nauk. SSSR **57** (1947), 657–659

[144] Schwartz, R., E., *Complex Hyperbolic Triangle Groups*, Proceedings of the International Congress of Mathematicians **1** (2002), 339–350

[145] Steinberg, S., *Lie Series, Lie Transformations, and their Applications*, in *Lie Methods in Optics*, Lec. Notes in Physics **250**, Springer V., Leon, Mexico (1985)

[146] Suzuki M., *Fractal Decomposition of Exponential Operators with Applications to Many-Body Theories and Monte Carlo Simulations*, Ph. Letters A **146** (1990), 319–323

[147] Suzuki, M., *General Theory of higher-order decomposition of exponential operators and symplectic integrators*, Physics Letters A **165**, North-Holland (1992), 387–395

[148] Suzuki, M., *General nonsymetric higher-order decompositions of exponential operators and symplectic integrators*, Physic Letters A **165** (1993), 387–395

[149] Toledo, D., *Representations of surface groups in complex hyperbolic space*, J. Diff. Geometry **29** (1989), 125–133

[150] Viennot, G., *Algèbres de Lie libres et Monoïdes Libres*, Lect. Notes in Math. **691**, Springer Verlag, 1978

[151] Wannier, G. H., *Antiferromagnetism. The Triangular Ising Net*, Phys. Rev. **79** (1950), 357–364

[152] Wisdom J., Holman, M., *Symplectic Maps for the N-Body Problem*, The Astr. J. **102(4)** (1991), 1528–1538

[153] YOSHIDA, H., *Conserved Quantities of Symplectic Integrators for Hamiltonian Systems*, Physica D, Springer V. (1990)

[154] H. YOSHIDA, *Construction Of Higher Order Symplectic Integrators*, Physics Letters A **150**, Elsevier Science Publishers B.V. (1990), 262–268

ANNEXES

Exhaustive Search of Symplectic Integrators Using Computer Algebra

KOSELEFF, P.-V.
Fields Institute Communications, 10 (1996)

Exhaustive Search of Symplectic Integrators using Computer Algebra *

P.-V. Koseleff

Équipe "Analyse Algébrique", Institut de Mathématiques
Université Pierre et Marie Curie, case 247
4 place Jussieu, F-75252 Paris Cedex 05
.e-mail : `koseleff@math.jussieu.fr`

Abstract. We find symplectic integrators using universal exponential identities or relations among formal Lie series. We give here general methods to compute such identities in a free Lie algebra. We recover by these methods all the previously known symplectic integrators and some new ones. We list all minimal solutions for integrators of low order. We give some improvement in the case when the Hamiltonian is in form $T(p)+V(q)$. We give also all reversible fourth-order symplectic integrators for the planetary hamiltonian expressed in canonical heliocentric coordinates.

1 Introduction

For very long time integration, there has been recently a development of numerical methods preserving the symplectic structure (see for example [7, 18, 19, 20]), which seem to be more efficient with respect to the computational cost.

Symplectic integrators may be seen as the time evolution mapping of a slightly perturbed Hamiltonian, that is to say as a Lie transformation that can be represented either by an exponential, a product of increasing order single exponentials or a proper Lie transformation. Constructing explicit high order symplectic integrators requires the manipulation of formal identities like exponential identities.

In section **2.**, we remind first some definitions of the Hamilton formalism. In section **3.**, we give some general methods to manipulate formal Lie series and Lie algebra automorphisms. We remind some theorems related to exponential identities and give explicit methods to compute them. They make use the Lyndon basis, which is particularly adapted to this problem. Most of this material has been published already in [9] but is not necessarily known by the reader.

In section **4.**, we show how the algorithms described in section **3.** provide symplectic integrators. The idea of such constructions originates in Forest & Ruth ([7]) or more recently Yoshida ([20]). Our approach in this paper is to

* Published in *Integration Algorithms and Classical Mechanics*, Fields Institut communications, **10**, pp. 103-120, 1996

combine the use of proper Lie transforms and exponentials. This avoids many unnecessary direct calculations of exponential identities. At the end we propose some improvement in the case when the Hamiltonian is separated into kinetic and potential energies. We give also some fourth-order integrator for the planetary Hamiltonian and show that they are minimals.

All the algorithms described in the present paper have been implemented using Axiom (NAG) running on IBM-RS/6000-550.

Between the preparation of this paper and its publication, many papers have been published on the subject and specially [13].

1.1 Symplectic Integrators

Given a phase space E which can be identified to $\mathbb{R}^{2n}$, a set of variables

$$(q,p) = (q_1,\ldots,q_n,p_1,\ldots,p_n) = (z_1,\ldots,z_{2n}), \tag{1}$$

and an Hamiltonian $h = h(p,q,t)$, we consider the system of differential equations

$$\dot{p}_i = -\frac{\partial h}{\partial q_i}, \dot{q}_i = \frac{\partial h}{\partial p_i}, \quad 1 \le i \le n, \tag{2}$$

where $\dot{z} = \frac{dz}{dt}$ denotes the total time derivative. Introducing the Poisson bracket

$$\{f,g\} = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i} - \frac{\partial g}{\partial p_i}\frac{\partial f}{\partial q_i} \tag{3}$$

that turns the set of smooth functions on E onto a Lie algebra, (2) becomes

$$\dot{z}_i = \{z_i,h\} = -L_h z_i, \quad 1 \le i \le 2n. \tag{4}$$

A transformation on the phase space E is said canonical if it preserves the Poisson brackets. Such transformations are also called symplectic as their Jacobians belong to the symplectic group. One extends the canonical transformations on the functions on the phase space by $Tf(z) = f(T(z))$. Canonical transformations act on the Lie algebra of the Lie operators by $TL_fT^{-1} = L_{Tf}$. Here $L_f : g \mapsto \{f,g\}$. The set of L_f is a Lie algebra with $[Lf, L_g] = L_{\{f,g\}}$.

The time-evolution mapping $S_h(t) : z \mapsto z(t)$ is a canonical transformation. From (4), $S_h(t)$ is the solution of the differential equation

$$\frac{d}{dt}S_h(t) = -S_h(t)L_h, S_h(0) = \mathbb{1}. \tag{5}$$

If h is not time-dependent we have $S_h(t) = e^{-tL_h}$ and a formal solution of (4) is given by its Taylor series, called Lie series

$$z(t) = \sum_{n\ge 0}(-t)^n \frac{L_h^n}{n!} z. \tag{6}$$

Published in Fields Institute Communications, **10**, 1996 3

If h is time-dependent, say for example $h = \sum_{n\geq 0} t^n h_n$, then $S_h(t)z$ may be written as a Lie series. If

$$Z = S_h(t)z = \sum_{n\geq 0} t^n Z_n,$$

we get from (4)

$$Z_0 = z, \quad Z_n = -\sum_{p=1}^{n} \tfrac{1}{n} L_{h_p} Z_{n-p}.$$

More generally, see ([3]), $S_h(t)$ is a series of operators $\sum_{n\geq 0} t^n (S_h)_n$ where

$$(S_h)_0 = \mathbb{1}, (S_h)_n = -\sum_{p=1}^{n} \tfrac{1}{n} L_{h_p} (S_h)_{n-p}.$$

⋆⋆⋆

Let us consider an Hamiltonian $h = A + B$, the two time-evolution mappings $S_A(t) = e^{-tL_A}$ and $S_B(t) = e^{-tL_B}$, and a given integer k, one seeks a minimal set of coefficients $c_1, \ldots, c_n, d_1, \ldots d_n$, such that

$$S^{(n)}(t) = S_A(c_1 t) S_B(d_1 t) \cdots S_A(c_n t) S_B(d_n t) = e^{-tL_h} + o(t^k). \tag{7}$$

$S^{(n)}(t)$ is a canonical transformation as composition of canonical transformations. The above expression may be considered as an equality between truncated Lie series. The aim of our paper is to show how one can solve this general problem considering the equation (7) as an universal identity between formal transformations on Lie algebra.

2 Lie algebraic formalism

In hamiltonian mechanics, the use of Lie methods or Lie transformations is efficient when it becomes easy to manipulate Lie polynomials and to express exponential identities like the Baker-Campbell-Hausdorff formula. Our aim in this section is to give general methods for the computation of such identities.

These identities are universal Lie algebraic identities, that is to say they do not depend on the Lie algebra we work in or the Lie bracket we use. We work in free Lie algebras and with formal Lie series, neglecting all the convergence problems that can appear with analytical functions for example.

We will use the Lyndon basis for the formal computations but all the identities can be later evaluated in any Lie algebra.

2.1 Definitions

X will denote an alphabet, that is to say an ordered set (possibly endless). R is a ring which contains the rational numbers $\mathbb{Q}$. X^* is the free monoid generated by X and is totally ordered with the lexicographic order. $M(X)$ is the free magma generated by X. It contains X and is equipped with a composition law : $(x, y) \mapsto (x, y)$.

$\mathcal{A}(X, R)$ is the associative algebra, that is to say the R-algebra of X^*.

A Lie algebra is an algebra in which the multiplication law $[,]$ is bilinear, alternate and satisfies the Jacobi identity:

$$[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0. \tag{8}$$

$L(X, R)$ or $L(X)$ is the free Lie algebra on X. It is defined as the quotient of the R-algebra of $M(X)$ by the ideal generated by the elements (u, u) and $(u, (v, w)) + (v, (w, u)) + (w, (u, v))$.

An element of $M(X)$ considered as element of $L(X)$ will be called a Lie monomial. $L_n(X)$ is the free module generated by those of length n. Thus $L(X)$ is graded by the length denoted by $|x|$ for $x \in M(X)$. If $|X| = q < \infty$ we have Witt's formula (see [1, 14, 15]):

$$\sum_{d|n} d \dim L_d(X) = q^n. \tag{9}$$

Given a weighted alphabet X in which each letter a has an integer weight $||a||$, we take the weight as graduation for $L(X)$ which is defined as the unique extension of the weight in X. We denote by $\tilde{L}_n(X)$ (resp. $\tilde{\mathcal{A}}_n(X)$) the submodule of $L(X)$ (resp. $\mathcal{A}(X)$) spanned by the elements of weight n.

2.2 Formal Lie series

We define the formal Lie series $\tilde{L}(X)$ and $\tilde{\mathcal{A}}(X)$ as

$$\tilde{L}(X) = \prod_{n \geq 0} \tilde{L}_n(X) \quad \text{and} \quad \tilde{\mathcal{A}}(X) = \prod_{n \geq 0} \tilde{\mathcal{A}}_n(X). \tag{10}$$

We will write $x \in \tilde{L}(X)$ as a series $\sum_{n \geq 0} x_n$. $\tilde{L}(X)$ is a complete Lie algebra with the Lie bracket

$$([x, y])_n = \sum_{p+q=n} [x_p, y_q]. \tag{11}$$

Denoting by $\tilde{L}(X)^+$ (resp. $\tilde{\mathcal{A}}(X)^+$) the ideal of $\tilde{L}(X)$ (resp. $\tilde{\mathcal{A}}(X)$) generated by the elements of positive weight, we can define the exponential and the logarithm as

$$\begin{array}{ll} \exp : \tilde{\mathcal{A}}(X)^+ \to 1 + \tilde{\mathcal{A}}(X)^+ & \log : 1 + \tilde{\mathcal{A}}(X)^+ \to \tilde{\mathcal{A}}(X)^+ \\ \qquad x \mapsto \sum_{n \geq 0} \dfrac{x^n}{n!} & \qquad x \mapsto -\sum_{n \geq 1} \dfrac{(1-x)^n}{n}. \end{array} \tag{12}$$

They are mutually reciprocal functions and we have (see [1, Ch. II, §5]) the

Theorem 1 (Campbell-Hausdorff). *If $x, y \in \tilde{L}(X)^+$ then*

$$\log\left[\exp(x)\exp(y)\right] \in \tilde{L}(X)^+. \tag{13}$$

Using the preceding lemma we deduce (see [4, 16]) the

Proposition 2 (Factored product expansion). *Given $k \in \tilde{L}(X)^+$, there is a unique series $g \in \tilde{L}(X)^+$ such that*

$$\exp(\textstyle\sum_{n\geq 1} k_n) = \cdots \exp(g_n) \cdots \exp(g_1). \tag{14}$$

2.3 Lie series automorphisms

For x in $\tilde{L}(X)$ we denote the Lie operator $L_x y = [x, y]$ by $L(x)$ or L_x. From the Jacobi identity (8) we have $[L_x, L_y] = L_x L_y - L_y L_x = L_{[x,y]}$. The set of L_x is a Lie algebra that we call the adjoint Lie algebra. For any Lie series automorphisms T, we have by definition $[Tf, Tg] = T[f, g]$. The Lie series automorphisms act on the adjoint Lie algebra by

$$T L_f T^{-1} = L_{Tf}. \tag{15}$$

Let us give now some example of Lie transformations that play an important role in hamiltonian mechanics.

The exponential. Given $x \in \tilde{L}(X)^+$, we consider $\exp(L_x)$ defined as

$$\exp(L_x) y = \sum_{i\geq 0} \frac{L_x^i}{i!} y. \tag{16}$$

From the Campbell-Hausdorff theorem (1), the set of all $\exp(L_x)$ is a group $\mathbf{G}$ that we will call the Lie transformations group.

The Lie transform. For $w = \sum_{n\geq 1} t^n w_n$, we denote by T_w and T_w^{-1} the solution ([3]) of

$$\frac{d}{dt} T_w = -T_w L_{\frac{dw}{dt}} \quad \text{and} \quad \frac{d}{dt} T_w^{-1} = L_{\frac{dw}{dt}} T_t^{-1}. \tag{17}$$

With the notation of (5), we have $T_w^{-1} = S_{\frac{dw}{dt}}$. For $g \in \tilde{L}(X)$, we have (see [2, 3]) $T_w^{-1} g = \sum_{n\geq 0} G_n$ where

$$G_0 = g_0, \quad G_{0,n} = g_n, \quad G_{p,q} = \sum_{k=1}^{p} \tfrac{k}{p}[w_k, G_{p-k,q}], \quad G_n = \sum_{p=0}^{n} G_{p,n-p}. \tag{18}$$

We call this transformation the Deprit transform. The composite T of two Lie transforms T_u and T_v satisfies

$$\frac{dT}{dt} = \frac{dT_u}{dt}T_v + T_u\frac{dT_v}{dt} = -T_u L_{\frac{du}{dt}} T_v - T_u T_v L_{\frac{dv}{dt}} \tag{19}$$

$$= -T\left(T_v^{-1} L_{\frac{du}{dt}} T_v + L_{\frac{dv}{dt}}\right) \tag{20}$$

$$= -T L_{T_v^{-1}\frac{du}{dt}+\frac{dv}{dt}}. \tag{21}$$

So $T = T_w$ where $\dfrac{dw}{dt} = T_v^{-1}\dfrac{du}{dt} + \dfrac{dv}{dt}$.

The Dragt-Finn transform. The Dragt-Finn transform is the infinite product of exponential maps (see [4]). Given $g = \sum_{n\geq 1} g_n$, we define M_g and M_g^{-1} as

$$M_g = \exp(-L_{g_1})\cdots\exp(-L_{g_n})\cdots, \quad M_g^{-1} = \cdots\exp(L_{g_n})\cdots\exp(L_{g_1}). \tag{22}$$

Note that this transformation will be used in this paper as a technical support for proving relations between Lie series.

2.4 Relations between transformations

The three above transformations are totally defined by generating series and are connected by the following

Proposition 3 ([8]). *Given $w, k, g \in \tilde{L}(X)^+$, there exist*

— $k' \in \tilde{L}(X)^+$ *with* $k'_n - w_n \in L(w_1, \ldots w_{n-1})$ *such that* $\exp(L_{k'}) = T_w^{-1}$,
— $g' \in \tilde{L}(X)^+$ *with* $g'_n - k_n \in L(k_1, \ldots k_{n-1})$ *such that* $M_{g'}^{-1} = \exp(L_k)$,
— $w' \in \tilde{L}(X)^+$ *with* $w'_n - g_n \in L(g_1, \ldots g_{n-1})$ *such that* $T_{w'}^{-1} = M_g^{-1}$.

The third part of the above proposition has been already proved by Finn ([6]), but not in terms of Lie polynomials. One proves (see [8, 10]) that $T_w = M_g$ if and only if

$$\frac{dw}{dt} = \sum_{n\geq 1} t^{n-1}\left[\sum_{k=1}^{n} k \sum_{\substack{(k+1)m_{k+1}+\cdots \\ +(n-k)m_{n-k}=n-k}} \frac{L_{g_{n-k}}^{m_{n-k}}\cdots L_{g_{k+1}}^{m_{k+1}}}{m_{k+1}!\cdots m_{n-k}!} g_k\right], \tag{23}$$

or equivalently

$$w_n = \sum_{k=1}^{n}\frac{k}{n} \sum_{\substack{(k+1)m_{k+1}+\cdots \\ +(n-k)m_{n-k}=n-k}} \frac{L_{g_{n-k}}^{m_{n-k}}\cdots L_{g_{k+1}}^{m_{k+1}}}{m_{k+1}!\cdots m_{n-k}!} g_k = g_n + G_n \tag{24}$$

in which $G_n \in L(g_1, \ldots, g_{n-1})$.

Using the proposition (2), one proves the existence of $g = \sum_{n\geq 1} g_n$ such that

$$\exp(\textstyle\sum_{n\geq 1} L_{k_n}) = \cdots \exp(L_{g_n}) \cdots \exp(L_{g_1}), \tag{25}$$

in which $g_n = k_n + K_n$ and $K_n \in L(k_1, \ldots, k_{n-1})$. Combining (24) and (25) we deduce proposition 3.

⁂

We deduce in passing that any Lie transformation $T \in \mathbf{G}$ may be expressed as an exponential of a Lie operator or as an infinite product of single exponentials or as a proper Lie transform. The use of a representation depends deeply on the result we look for. For example, if we have to compose transformations, it is much easier to consider Lie transforms because their product is a Lie transform whose generating function appears easily.

Explicit relations up to any given order may be easily computed, using the Lyndon basis. For example, given $w = \sum_{n\geq 1} w_n$, we have at the order 6 $\exp(L_k) = T_w^{-1}$ in which

$$\begin{aligned}
k = {} & w_1 + w_2 + w_3 - \tfrac{1}{6}\,[w_1,\, w_2] + w_4 - \tfrac{1}{4}\,[w_1,\, w_3] + w_5 - \tfrac{3}{10}\,[w_1,\, w_4] \\
& - \tfrac{1}{10}\,[w_2,\, w_3] + \tfrac{1}{120}\,[w_1,\, [w_1,\, w_3]] + \tfrac{1}{60}\,[[w_1,\, w_2],\, w_2] \\
& + \tfrac{1}{360}\,[w_1,\, [w_1,\, [w_1,\, w_2]]] + w_6 - \tfrac{1}{3}\,[w_1,\, w_5] - \tfrac{1}{6}\,[w_2,\, w_4] \\
& + \tfrac{1}{60}\,[w_1,\, [w_1,\, w_4]] + \tfrac{1}{30}\,[w_1,\, [w_2,\, w_3]] + \tfrac{1}{24}\,[[w_1,\, w_3],\, w_2] \\
& + \tfrac{1}{240}\,[w_1,\, [w_1,\, [w_1,\, w_3]]] - \tfrac{1}{180}\,[w_1,\, [[w_1,\, w_2],\, w_2]]
\end{aligned}$$

3 Search of symplectic integrators

Let $h = A + B$ be an Hamiltonian. For given integers n and k, one looks for a set $c_1, \ldots, c_n, d_1, \ldots d_n$, such that

$$S^{(n)}(t) = S_A(c_1 t) S_B(d_1 t) \cdots S_A(c_n t) S_B(d_n t) = S_h(t) + o(t^k). \tag{26}$$

Let $R = \mathbb{Q}[c_1, \ldots, c_n, d_1, \ldots, d_n]$, $\tilde{L}_p(\{A, B\})$ be the submodule of $\tilde{L}(\{A, B\})$ spanned by elements of weight p, where $||A|| = ||B|| = 1$. Using the Witt formula (9), the dimension l_p of $\tilde{L}_p(\{A, B\})$ satisfies $\sum_{d|n} d l_d = 2^n$.

Let $\tilde{L}_p\{z, A, B\}$ be the submodule of $\tilde{L}(\{z, A, B\})$, spanned by those of weight p, in which the partial commutative degree in z is 1 (here $||z|| = 0$). Using the Lyndon basis, one proves directly that the dimension of $\tilde{L}_p\{z, A, B\}$ is 2^p.

⁂

In order to solve equation (26), we have to express exponential identities first. We first prove the algebraic equalities obtained by representing the integrators in several ways.

Proposition 4. *Let us consider the following problems:*

1. *Having expressed* $S^{(n)}(t)$ *as* $\exp(-\sum_{p\geq 1} t^p L_{K_p})$*, solve* $K_1 = A+B, K_2 = \cdots = K_k = 0$,
2. *Having expressed* $S^{(n)}(t)$ *as* $\exp(-tL_{G_1})\cdots\exp(-t^p L_{G_p})\cdots$*, solve* $G_1 = A+B, G_2 = \cdots = G_k = 0$,
3. *Having expressed* $S^{(n)}(t)$ *as* $T_{(tW_1+\cdots+t^pW_p+\cdots)}$*, solve* $W_1 = A+B, W_2 = \cdots = W_k = 0$.
4. *Having expressed* $Z = S^{(n)}(t)z$*, solve* $Z_0 = (\exp(-tL_{A+B})z)_0, \ldots, Z_k = (\exp(-tL_{A+B})z)_k$.

The solutions of these four problems are the zeroes of the same polynomial ideal.

Denoting the Lyndon basis $\mathcal{L}_p(\{A,B\})$ by $(x_{p,1},\ldots,x_{p,l_p})$, we have

$$K_p = \sum_{i=1}^{l_p} K_{p,i}x_{p,i}, G_p = \sum_{i=1}^{l_p} G_{p,i}x_{p,i}, W_p = \sum_{i=1}^{l_p} W_{p,i}x_{p,i}. \tag{27}$$

For the first three methods, the solutions are the zeroes of the ideals

$$\begin{aligned} \mathcal{I}_K^{(k)} &= (K_{1,1}-1, K_{1,2}-1) + (K_{i,j}; 2\leq i\leq k, 1\leq j\leq l_i),\\ \mathcal{I}_G^{(k)} &= (G_{1,1}-1, G_{1,2}-1) + (G_{i,j}; 2\leq i\leq k, 1\leq j\leq l_i),\\ \mathcal{I}_W^{(k)} &= (W_{1,1}-1, W_{1,2}-1) + (W_{i,j}; 2\leq i\leq k, 1\leq j\leq l_i). \end{aligned} \tag{28}$$

We have to bear in mind that S_W is the Deprit transform associated to $\int_0^t W(u)\,du$. We therefore get the relations due to the proposition 3:

$$\begin{aligned} K_p &= \tfrac{1}{p}W_p + R_p^W, \quad R_p^W \in L_p(W_1,\ldots,W_{p-1}),\\ \tfrac{1}{p}W_p &= G_p + R_p^G, \quad R_p^G \in L_p(G_1,\ldots,G_{p-1}),\\ G_p &= K_p + R_p^K, \quad R_p^K \in L_p(K_1,\ldots,K_{p-1}). \end{aligned} \tag{29}$$

We first have $K_1 = G_1 = W_1$. For $p > 1$, each R_p^W may be expressed in the Lyndon basis $\mathcal{L}_p(\{A,B\})$ and the coefficients are polynomials. Each monomial contains a $W_{i,j}$ where $i > 1$ and thus belongs to $\mathcal{I}_W^{(k)}$. We therefore deduce that $\mathcal{I}_K^{(k)} \subset \mathcal{I}_W^{(k)}$. On the same way, we deduce $\mathcal{I}_W^{(k)} \subset \mathcal{I}_G^{(k)}$ and $\mathcal{I}_G^{(k)} \subset \mathcal{I}_K^{(k)}$.

We thus have $\mathcal{I}_K^{(k)} = \mathcal{I}_W^{(k)} = \mathcal{I}_G^{(k)}$.

⁂

Z_p belongs to $\tilde{L}_p\{z,A,B\}$ which basis is denoted by $z_{p,1},\ldots,z_{p,2^p}$. If $Z_p = \sum_{q=1}^{2^p} Z_{p,q}z_{p,q}$ and $(\exp(-tL_{A+B})z)_p = \sum_{q=1}^{2^p} a_{p,q}z_{p,q}$, the solutions of **4.** are the zeroes of

$$\mathcal{I}_Z^{(k)} = (Z_{i,j} - a_{i,j}, 1\leq i\leq k, 1\leq j\leq 2^i) \tag{30}$$

From $S^{(n)}(t)z - \exp(-tL_{A+B})z = \left(S^{(n)}(t) - \exp(-tL_{A+B})\right)z = o(t^k)$, we deduce for $p=1$

$$Z_1 - \{z, A+B\} = -(L_{G_1} - L_{A+B})z = \{z, G_1 - (A+B)\} = 0, \qquad (31)$$

that is, onto the basis $([z,A],[z,B])$,

$$Z_{1,1} - 1 = G_{1,1} - 1 = 0, \quad Z_{1,2} - 1 = G_{1,2} - 1 = 0. \qquad (32)$$

For $1 < p \leq k$, we have

$$\begin{aligned} Z_p - \tfrac{(-1)^p}{p!} L^p_{A+B} z &= \left[\sum_{m_1 + \cdots + p m_p = p} (-1)^{m_1 + \cdots + m_p} \frac{L^{m_1}_{G_1} \cdots L^{m_p}_{G_p}}{m_p! \cdots m_1!} - (-1)^p \frac{L^p_{A+B}}{p!} \right] z \\ &= \tfrac{(-1)^p}{p!} \left(L^p_{G_1} - L^p_{A+B}\right) z - L_{G_p} z + & (33) \\ &\qquad \sum_{\substack{m_1 + \cdots + (p-1) m_{p-1} = p \\ m_1 < p}} (-1)^{m_1 + \cdots + m_{p-1}} \frac{L^{m_1}_{G_1} \cdots L^{m_{p-1}}_{G_{p-1}}}{m_p! \cdots m_1!} z & (34) \\ &= 0. & (35) \end{aligned}$$

$L^p_{G_1} - L^p_{A+B}$ may be written as

$$\sum_{k_1 + \cdots + k_p + l_1 + \cdots + l_p = p} \left(G^{k_1 + \cdots + k_p}_{1,1} G^{l_1 + \cdots + l_p}_{1,2} - 1 \right) L^{k_1}_A L^{l_1}_B \cdots L^{k_p}_A L^{l_p}_B. \qquad (36)$$

A coefficient of (36) may be expressed as

$$G^k_{1,1} G^l_{1,2} - 1 = (G^k_{1,1} - 1) G^l_{1,2} + (G^l_{1,2} - 1) \in (G_{1,1} - 1, G_{1,2} - 1). \qquad (37)$$

Each $L^{k_1}_A L^{l_1}_B \cdots L^{k_p}_A L^{l_p}_B$ is a sum of $z_{p,q}$ so any coefficient in $(L^p_{G_1} - L^p_{A+B})z$ belongs to $(G_{1,1}-1, G_{1,2}-1) \subset \mathcal{I}^{(p)}_G$. Other terms in (33) and (34) have coefficients in $(G_{2,1}, \ldots, G_{p,1}, \ldots G_{p,l_p}) \subset \mathcal{I}^{(p)}_G$. We thus deduce that $\mathcal{I}_Z(p) \subset \mathcal{I}^p_G$.

⁂

Suppose we have for each $m < p$, $\mathcal{I}^{(m)}_Z = \mathcal{I}^{(m)}_G$. At the order p, from (35), we deduce also that

$$\begin{aligned} -L_{G_p} z = \overbrace{Z_p - \tfrac{(-1)^p}{p!} L^p_{A+B} z}^{(I)} - \overbrace{\tfrac{(-1)^p}{p!} \left(L^p_{G_1} - L^p_{A+B}\right) z}^{(II)} - \\ \underbrace{\sum_{m_1 + \cdots + m_{p-1} = p, m_1 < n} (-1)^{m_1 + \cdots + m_{p-1}} \frac{L^{m_1}_{G_1} \cdots L^{m_{p-1}}_{G_{p-1}}}{m_p! \cdots m_1!} z}_{(III)}. \qquad (38) \end{aligned}$$

Each term on the r.h.s. of the above equation, has a decomposition onto the Lyndon basis $z_{p,q}$. Coefficients of (I) belong to $\mathcal{I}_Z^{(p)}$, coefficients of (II) and (III) belong to $(G_{1,1}-1, G_{1,2}-1) \subset \mathcal{I}_G^{(p-1)} \subset \mathcal{I}_Z^{(p-1)} \subset \mathcal{I}_Z^{(p)}$.

We deduce that the coefficients of the $L_{G_p} z$ decomposition onto the $z_{p,q}$ belong to $\mathcal{I}_Z^{(p)}$. As solution of (38), the coefficients of G_p onto the $x_{p,q}$ belong to $\mathcal{I}_Z^{(p)}$ We thus have $\mathcal{I}_G^{(p)} \subset \mathcal{I}_Z^{(p)}$ and eventually $\mathcal{I}_G^{(p)} = \mathcal{I}_Z^{(p)}$.

⁂

In order to obtain symplectic integrators, we can use one of these methods which are algebraically equivalent. We will not use the Dragt-Finn representation as its mathematical interpretation in terms of invariants is not clear. Nevertheless, composition of Factored Product transformations is widely used in optics ([5]).

Direct method. The problem (26) may be solved, looking for all z

$$S^{(n)}(t)z = S_A(c_1 t)S_B(d_1 t)\cdots S_A(c_n t)S_B(d_n t)z = S_h(t)z + o(t^k). \tag{39}$$

At each order p, we obtain a system of 2^p polynomial equations.

Invariant function. Problem (39) may be solved by expressing $S^{(n)}$ as an exponential and looking for $c_1, \ldots, c_n, d_1, \ldots d_n$ such that

$$S^{(n)}(t) = S_A(c_1 t)S_B(d_1 t)\cdots S_A(c_n t)S_B(d_n t) = e^{-tL_K} \tag{40}$$

with $K = h + o(t^{k-1})$. K is not the Hamiltonian governing the system given by $S^{(n)}$, but an invariant function. At each order p, there is l_p polynomial equations to solve. In order to get these, we have to compute some Baker-Campbell-Hausdorff formulas.

Perturbed Hamiltonian. One can also express $S^{(n)}(t)$ as a Lie transform

$$S_A(c_1 t)S_B(d_1 t)\cdots S_A(c_n t)S_B(d_n t) = S_W \tag{41}$$

where $W = h + o(t^{k-1})$. The condition is obtained by writing

$$\frac{d}{dt}S_W = -S_W L_W = -S_h L_h + o(t^{k-1}). \tag{42}$$

W is the Hamiltonian governing the system which time-evolution mapping is $S^{(n)}$. At each order p, there is the same number of equations as previously but that avoids many unnecessary direct calculations of Baker-Campbell-Hausdorff formulas.

Published in Fields Institute Communications, **10**, 1996

3.1 First integrators

For a given order k, and for a given method (see table 1) one seeks a minimal n such that $S^{(n)}(t)$ is a kth-order symplectic integrator. One looks for the zeroes of a polynomial ideal. We use here algebraic methods like Gröbner basis that we compute using Axiom (when possible) or Macaulay that works in a ring $\mathbb{Z}/p\mathbb{Z}$ and gives some precious results.

For low orders, these methods furnish symplectic integrators. All the following results have been obtained with algorithms on Lie series and have been implemented t in Axiom ([8]). We then obtain the polynomials that define the variety we look at. For low orders, these can be described ([9]).

• The solution for $k = 1$ is given by $c_1 = d_1 = 1$ and

$$S_1(t) = S^{(2)}(t) = S_A(t)S_B(t). \tag{43}$$

• The solution for $k = 2$ is given by

$$S_2(t) = S^{(3)}(t) = S_A(\tfrac{t}{2})S_B(t)S_A(\tfrac{t}{2}) = S_1(t)S_1^{-1}(-t). \tag{44}$$

This approximant is reversible, that is to say satisfies $S_2^{-1}(t) = S_2(-t)$.

• Solutions for $k = 3$ are

$$\begin{aligned} S_3(t) &= S_A(ct)S_B(ct)S_B(ct)S_A(ct)S_A(\bar{c}t)S_B(\bar{c}t)S_B(\bar{c}t)S_A(\bar{c}t) \\ &= S_2(ct)S_2(\bar{c}t) \end{aligned} \tag{45}$$

where $c^2 - \frac{1}{2}\,c + \frac{1}{12} = 0$.

• When $k = 4$, we find two sets of solutions and 5 solutions. The corresponding integrators have 7 factors. If $c^3 - 2c^2 + c = \frac{1}{6}$, we get

$$S_4(t) = S_2(ct)S_2((1-2c)t)S_2(ct). \tag{46}$$

These solutions are known and have been given also by Yoshida ([20]).

If c is a root of $c^2 - \frac{1}{2}c + \frac{1}{6} = 0$, we get two other solutions:

$$S_4(t) = S_2(ct)S_2((c+\bar{c})t)S_2(\bar{c}t). \tag{47}$$

• For $k = 5$, one gets exactly 46 solutions as product of 5 approximants S_2.

• For $k = 6$, an exhaustive list is still unknown. If we add the condition of being reversible, we get at most 39 solutions, working in $\mathbb{Z}/31991\mathbb{Z}$. All integrators are products of second-order integrators. The real valued integrators for $k = 2$ or 4 are reversible, that means $S(-t) = S^{-1}(t)$.

3.2 Reversible Integrators

Representing a reversible integrator $S(t)$ by an exponential $\exp(-tL_K)$, we deduce that $K(t) = K(-t)$. Looking for reversible integrators, we can deduce from the Campbell-Hausdorff formula the

Lemma 5 ([17]). *If $S_{2k}(t)$ is a reversible symplectic integrator of order $2k$, then*

$$S(t) = S_{2k}\left(\frac{1}{2-\sqrt[2k+1]{2}}t\right) S_{2k}\left(-\frac{\sqrt[2k+1]{2}}{2-\sqrt[2k+1]{2}}t\right) S_{2k}\left(\frac{1}{2-\sqrt[2k+1]{2}}t\right)$$

is a reversible symplectic integrator of order $2k+2$.

This lemma allows us to build reversible symplectic integrators of order $2k$ as products of $2.3^{k-1}+1$ single operators S_A or S_B. With this method we find a 19-factor sixth-order integrator.

⁂

One can try to find directly reversible integrators looking for reversible products

$$S_R^{(n)}(t) = S_A(c_n t)S_B(d_n t)\cdots S_A(c_1 t)S_B(d_1 t)S_A(c_0 t)S_B(d_1 t)S_A(c_1 t)\cdots S_B(d_n t)S_A(c_n t).$$

Denoting by $S^{(n)}(t)$ the operator

$$S_A(c_n t)S_B(d_n t)\cdots S_A(c_1 t)S_B(d_1 t)S_A(\tfrac{c_0}{2}t), \tag{48}$$

we obtain $S_R^{(n)}$ as $S^{(n)}(t){S^{(n)}}^{-1}(-t)$ that we can express as an exponential or a Lie transform. Representing $S_R^{(n)}$ as an exponential e^{-tL_K} has the advantage that $K(t) = K(-t)$. Moreover we have the following lemma resulting from proposition 3:

Lemma 6. *Let $\mathcal{I}_W^{(k)} = (W_{1,1}-1, W_{1,2}-1, W_{2,1}, \ldots, W_{k,1}, \ldots, W_{k,l_k})$, be the polynomial ideal defining the solutions of $S_r^{(n)}(t) = S_{W(t)} = S_h + o(t^k)$. For each $2p \le k$, we have*

$$\{W_{2p,1}, \ldots, W_{2p,l_{2p}}\} \subset \mathcal{I}_W^{(2p-1)} \quad \textit{so} \quad \mathcal{I}_W^{(2p)} = \mathcal{I}_W^{(2p-1)}. \tag{49}$$

Using proposition 4, we get for each $1 \le 2p \le k$,

$$\mathcal{I}_K^{(p)} = \mathcal{I}_W^{(p)} = (K_{1,1}-1, K_{1,2}-1, K_{2,1}, \ldots, K_{p,1}, K_{p,l_p}) \tag{50}$$

where the $K_{i,j}$ are the coefficient of K satisfying $\exp(-tK(t)) = S_r^{(n)}$. As S_W is reversible, we have $K_{2p} = 0$, for each p, that is $\mathcal{I}_K^{(2p)} = \mathcal{I}_K^{(2p-1)}$ and $\mathcal{I}_W^{(2p)} = \mathcal{I}_W^{(2p-1)}$.

⁂

This lemma proves that there is no need to consider odd terms of the Hamiltonian obtained with reversible integrators.

• For $k = 4$, one finds 3 reversible integrators obtained with the direct method.

• For $k = 6$, one proves that there is no solutions for $n < 8$. For $n = 8$, one sees, using the Hilbert function implemented in MACAULAY, that the variety of solutions in $\mathbb{Z}/p\mathbb{Z}$ ($p = 31991$) is constituted of 39 points. There is at most 39 algebraic solutions over $\mathbb{Q}$.

⁂

Another solution has been proposed by Yoshida [20] consisting in the finding of reversible integrators as reversible product of second-order integrators S_2. We look for

$$S^{(n)}(t) = S_2(c_n t) \cdots S_2(c_1 t) S_2(c_0 t) S_2(c_1 t) \cdots S_2(c_n t) = e^{-tL_{K^{(n)}}}. \qquad (51)$$

• For $k = 4$, we find the real valued reversible integrator previously found by the direct method or using the lemma (5).

• For $k = 6$, we have four equations with four unknowns $c_0, \ldots, c_3$. The solution is obtained after eliminations with

$$\begin{aligned} P_0(c_0) = {} & c_0^{39} + 4\ c_0^{38} - 18\ c_0^{37} - \frac{232}{3}\ c_0^{36} + \frac{6469}{45}\ c_0^{35} + \frac{8108}{15}\ c_0^{34} - \frac{82144}{135}\ c_0^{33} - \\ & \frac{239008}{135}\ c_0^{32} + \frac{870652}{675}\ c_0^{31} + \frac{5898416}{2025}\ c_0^{30} - \frac{618824}{675}\ c_0^{29} - \frac{5158016}{2025}\ c_0^{28} + \\ & \frac{2525372}{30375}\ c_0^{27} + \frac{32135888}{30375}\ c_0^{26} - \frac{1377776}{10125}\ c_0^{25} - \frac{33361568}{91125}\ c_0^{24} + \frac{536566}{10125}\ c_0^{23} + \\ & \frac{35651416}{455625}\ c_0^{22} - \frac{19660868}{1366875}\ c_0^{21} - \frac{8051504}{455625}\ c_0^{20} + \frac{5636474}{1366875}\ c_0^{19} + \frac{11313208}{4100625}\ c_0^{18} - \\ & \frac{17674448}{20503125}\ c_0^{17} - \frac{8733536}{20503125}\ c_0^{16} + \frac{1302268}{6834375}\ c_0^{15} + \frac{87632}{2460375}\ c_0^{14} - \frac{624184}{20503125}\ c_0^{13} + \\ & \frac{288448}{922640625}\ c_0^{12} + \frac{3333844}{922640625}\ c_0^{11} - \frac{716752}{922640625}\ c_0^{10} - \frac{127664}{553584375}\ c_0^{9} + \frac{143264}{922640625}\ c_0^{8} - \\ & \frac{136499}{4613203125}\ c_0^{7} - \frac{19996}{8303765625}\ c_0^{6} + \frac{117142}{41518828125}\ c_0^{5} - \frac{33848}{41518828125}\ c_0^{4} + \\ & \frac{17431}{124556484375}\ c_0^{3} - \frac{9668}{622782421875}\ c_0^{2} + \frac{656}{622782421875}\ c_0 - \frac{64}{1868347265625} = 0, \end{aligned}$$

and $c_1 = P_1(c_0), c_2 = P_2(c_0), c_3 = P_3(c_0)$ where P_1, P_2, P_3 are polynomials of degree 38. P_0 is irreducible over $\mathbb{Q}$ and has only three real roots. All the solutions are reached with this method as there is at most 39 solutions.

• For $k = 8$, Yoshida ([20]) has found 5 real valued integrators using numerical methods. These integrators involve 31 single integrators S_A or S_B. We proved, using standard basis computed with Macaulay, that these integrators are not products of 5 fourth-order symplectic integrators.

3.3 Special cases

Most of the times, when $h = T(p) + V(q)$, the kinetic energy is just a quadratic form in p. That means that $\{T, V\}$ is of degree one in p, $\{\{T, V\}, V\}$ depends only on q and $\{\{\{T,V\},V\},V\} = 0$. We may find symplectic integrators of order 4 or 6 involving less terms.

Unfortunately, there is no integrator of order 4 using less than 7 terms.

As $\{\{T,V\},V\}$ depends only on q, $V_1 = \alpha V + t^2\beta\{\{T,V\},V\}$ depends only on q and t for any α, β and we have

$$e^{-tL_{V_1}}p = p - t\frac{\partial V_1}{\partial q} \quad \text{and} \quad e^{-tL_{V_1}}q = q. \tag{52}$$

Denoting $e^{-t(\alpha L_V + \beta t^2 L_{\{\{T,V\},V\}})}$ by $S_{\alpha,\beta}(t)$ we look now for integrators $S^{(n)}$ as product of

$$S_{c_n,z_n}(t)S_T(d_nt)\cdots S_{c_1,z_1}(t)S_T(d_0t)S_{c_1,z_1}(t)\cdots S_T(d_nt)S_{c_n,z_n}(t) \tag{53}$$

or

$$S_T(d_nt)S_{c_n,z_n}(t)\cdots S_T(d_1t)S_{c_0,z_0}(t)S_T(d_1t)\cdots S_{c_n,z_n}(t)S_T(d_nt) \tag{54}$$

With this method we found a 5-factor fourth-order integrator and an 9-factor sixth-order integrator (see [9]).

3.4 Decomposition into more terms

One can generalize the search of symplectic integrators to the case of 3 operators (see for example Suzuki [17]).

In this part, we show how to find minimal symplectic integrators in the case when $h = A_1 + A_2 + A_3$, using the algorithms previously described. Such integrators can be useful for planetary problems written in the canonical heliocentric variables of Poincaré (see [12]).

We will give an exhaustive list for orders 1, 2 and 4. It is clear that any permutation on A_1, A_2, A_3 will also give an integrator.

The first-order integrator is

$$S_1(t) = S_{A_1}(t)S_{A_2}(t)S_{A_3}(t). \tag{55}$$

• From $S_2(t) = S_{A_2}(\frac{t}{2})S_{A_3}(t)S_{A_2}(\frac{t}{2})$, we deduce an second-order integrator for $A_1 + (A_2 + A_3)$:

$$\begin{aligned} S_2(t) &= S_{A_1}(\tfrac{t}{2})S_{A_2}(\tfrac{t}{2})S_{A_3}(t)S_{A_2}(\tfrac{t}{2})S_{A_1}(\tfrac{t}{2}). \\ &= S_1(\tfrac{t}{2})S_1^{-1}(-\tfrac{t}{2}) \end{aligned} \tag{56}$$

• Looking for reversible fourth-order integrators, we could deduce from the 7-factor fourth-order integrator S_4 for $A_1 + A_3$, that

$$S_{A_3}(d_2t)S_4(c_1t)S_{A_3}(d_1t)S_4(c_0t)S_{A_3}(d_1t)S_4(c_1t)S_{A_3}(d_2t) \tag{57}$$

is a 25-factor fourth-order integrator for $h = (A_1 + A_2) + A_3$.

Using the lemma 5, one obtains a 13-factor fourth-order symplectic integrator as a product of 3 reversible second-order integrators:

$$S_{A_1}(\tfrac{a}{2}t)S_{A_2}(\tfrac{a}{2}t)S_{A_3}(at)S_{A_2}(\tfrac{a}{2}t)S_{A_1}(\tfrac{a+b}{2}t)S_{A_2}(\tfrac{b}{2}t)S_{A_3}(bt)S_{A_2}(\tfrac{b}{2}t)$$
$$S_{A_1}(\tfrac{a+b}{2}t)S_{A_2}(\tfrac{a}{2}t)S_{A_3}(at)S_{A_2}(\tfrac{a}{2}t)S_{A_1}(\tfrac{a}{2}t). \qquad (58)$$

where $a = \frac{1}{2-\sqrt[3]{2}}$, $b = -\frac{\sqrt[3]{2}}{2-\sqrt[3]{2}}$.

Let us prove now that these integrators are minimal regards to the number of factors. For a given order k, a kth-order integrator is an operator of length m:

$$S^{(m)}(t) = S_{A_1}(x_1t)S_{A_2}(x_2t)S_{A_3}(x_3t)\cdots S_{A_m}(x_mt), \qquad (59)$$

where $i-1 = m \bmod 3$. A p-factor kth-order integrator is a sequence $(x_1,\dots,x_m)$ in which $x_1x_m \neq 0$ and $m-p$ of the x_i's are equals to zero. In such a sequence, there is no 2 consecutive zeroes, otherwise its length would be $m-2$. For the integrator (56), we have

$$x_1 = \tfrac{1}{2}, x_2 = \tfrac{1}{2}, x_3 = 1, x_4 = 0, x_5 = \tfrac{1}{2}, x_6 = 0, x_7 = \tfrac{1}{2}.$$

For given m and k the set of sequence $(x_1,\dots,x_m)$ satisfying

$$S_{A_1}(x_1t)S_{A_2}(x_2t)S_{A_3}(x_3t)\cdots S_{A_m}(x_mt) = S_h(t) + o(t^k), \qquad (60)$$

is an algebraic variety. For a given order k, let us denote by M_k the minimal integer such that each minimal kth-order integrator (up to a permutation of (A_1,A_2,A_3)) may be written as $S^{(m)}$ where $m \leq M_k$. Each kth-order integrator (up to a permutation of A_1,A_2,A_3) will be a sequence that is solution of

$$S^{(m)}(x_1,\dots,x_m) = S_{A_1}(x_1t)\cdots S_{A_m}(x_mt) = S_{A_1+A_2+A_3}(t) + o(t^k), \qquad (61)$$

with $m \leq M_k$.

The second-order example shows that $M_2 \geq 7$. Let S_2 be any 5-factor second-order integrator. Its length is at most 9. If its length is 9 then we have $x_2 = x_4 = x_6 = x_8 = 0$. By transposing A_2 and A_3, the sequence (x_1,x_3,x_5,x_7,x_9) gives also a 5-factor integrator.

If its length is 8, then there are 3 zeroes in the subsequence $x_2,\dots,x_7$, because $x_1x_8 \neq 0$. There is only two possibilities: $x_2 = x_4 = x_6 = 0$ or $x_3 = x_5 = x_7 = 0$. In the first case, by transposing A_2 and A_3, the sequence $(x_1,x_3,x_5,x_7,0,x_8)$ gives a 6-factor integrator. In the second case, the same transposition gives also a 5-factor integrator with the sequence $(x_1,0,x_2,x_4,x_6,x_8)$. It proves that any 5-factor operator (e.g. second-order integrator) (up to permutations) has a length less than 7.

- There is no 5-factor second-order integrator of length 5.
- Looking for integrators of length 6, one finds

$$S_2(t) = S_{A_1}(ct)S_{A_2}(ct)S_{A_3}(ct)S_{A_1}(\bar{c}t)S_{A_2}(\bar{c}t)S_{A_3}(\bar{c}t) = S_1(ct)S_1(\bar{c}t) \qquad (62)$$

where c is a complex root of $c^2 - c + \frac{1}{2}$.

• Looking for all possible solution involving 7 variables, we have $x_1x_7 \neq 0$ so for a 5-factor integrator we must have $x_2x_3x_5x_6 = 0$. We thus find

$$x_1 = x_7 = \tfrac{1}{2}, x_4 = 0, (x_2 - \tfrac{1}{2})x_2 = 0, x_2 = x_3 - \tfrac{1}{2} = \tfrac{1}{2} - x_6 = 1 - x_5. \quad (63)$$

If $x_2 = 0$ or $x_2 = 1$, we find the integrator (56). That proves that any minimal real second-order integrator has 5 factors and length 7.

⁂

Let S be a 7-factor operator. One can suppose that the 2 first factors are S_{A_1} and S_{A_2}. Let m be its length and express S as

$$S(t) = S_{A_1}(x_1t)S_{A_2}(x_2t)S_{A_3}(x_3t)\cdots S_{A_m}(x_mt) \quad (64)$$

in which $x_1x_2x_m \neq 0$. The subsequence $(x_2, \ldots x_{m-1})$ has a length $m-2$ and corresponds to a 5-factor operator. We thus deduce that $m-2 \leq 8$ because x_{m-1} could be zero. So any 7-factor operator may be written as an operator of length $m \leq 10$.

Suppose now that S has length 10 and that $x_1x_2x_{10} \neq 0$. There are 3 zeroes in the subsequence $x_3, \ldots, x_9$ and there are not consecutive. The only solutions are

a) $x_3 = x_5 = x_7 = 0$, *b*) $x_3 = x_5 = x_8 = 0$, *c*) $x_3 = x_5 = x_9 = 0$,
d) $x_3 = x_6 = x_8 = 0$, *e*) $x_3 = x_6 = x_9 = 0$, *f*) $x_3 = x_7 = x_9 = 0$,
g) $x_4 = x_6 = x_8 = 0$, *h*) $x_4 = x_6 = x_9 = 0$, *i*) $x_4 = x_7 = x_9 = 0$,
j) $x_5 = x_7 = x_9 = 0$.

Let us suppose that $S(t)S^{-1}(-t)$ is a reversible fourth-order integrator. If we suppose now that $A_1 = 0$, then $S(t)S^{-1}(t)$ is still a fourth-order integrator for $A_2 + A_3$. It implies that we must have at least 2 factors S_{A_2} and 2 factors S_{A_3} in S. So the only cases to consider are $a), g), j)$.

In the first case, suppose that $A_2 = 0$, then we obtain

$$S'(t) = S_{A_1}((x_1 + x_4)t)S_{A_3}((x_6 + x_9)t)S_{A_1}(x_{10}t).$$

In the second case, suppose that $A_3 = 0$, then we get

$$S'(t) = S_{A_1}(x_1t)S_{A_2}((x_2 + x_5)t)S_{A_1}((x_7 + x_{10})t).$$

In the third case, suppose that $A_1 = 0$, then we obtain

$$S'(t) = S_{A_2}(x_2t)S_{A_3}((x_3 + x_6)t)S_{A_2}(x_8t).$$

One of those case would imply that there is 5-factor fourth-order integrator which is impossible.

It shows that any minimal reversible fourth-order integrator S_4 may be found by looking for a 7 factor operator $S(t)$ of maximal length 9, such that $S_4 = S(t)S^{-1}(-t)$.

Published in Fields Institute Communications, **10**, 1996

⁂

If we look for all possible reversible 13 factor fourth-order integrators we get a zero dimensional algebraic variety of degree 12 and c_0 satisfies

$$(c_0^9 - 2\ c_0^8 + c_0^7 + \tfrac{2}{3}\ c_0^6 - c_0^5 + \tfrac{2}{3}\ c_0^3 - \tfrac{5}{9}\ c_0^2 + \tfrac{2}{9}\ c_0 - \tfrac{1}{27})(c_0^3 + c_0^2 - c_0 + \tfrac{1}{3}) = 0.$$

We therefore get 2 sets of solutions:

$$\begin{aligned}
&c_0 = 2e_1 = 2d_1,\\
&c_1 = e_2 = d_2 = 27\ c_0^8 - \tfrac{81}{2}\ c_0^7 + \tfrac{9}{2}\ c_0^6 + \tfrac{45}{2}\ c_0^5 - 15\ c_0^4 - 9\ c_0^3 + \tfrac{27}{2}\ c_0^2 - \tfrac{15}{2}\ c_0 + 2,\\
&c_2 = d_3 = e_3 = -c_1 - \tfrac{1}{2}c_0 - \tfrac{1}{2},\\
&c_0^9 - 2\ c_0^8 + c_0^7 + \tfrac{2}{3}\ c_0^6 - c_0^5 + \tfrac{2}{3}\ c_0^3 - \tfrac{5}{9}\ c_0^2 + \tfrac{2}{9}\ c_0 - \tfrac{1}{27} = 0.
\end{aligned}$$

and

$$\begin{aligned}
&c_1 = e_2 = 0,\\
&d_2 = d_3 = e_3 = \tfrac{1}{2}c_2 = -\tfrac{1}{4}\ c_0 + \tfrac{1}{4},\\
&d_1 = \tfrac{1}{2}c_0,\\
&e_1 = \tfrac{1}{4}c_0 + \tfrac{1}{4},\\
&c_0^3 + c_0^2 - c_0 + \tfrac{1}{3} = 0.
\end{aligned}$$

The first set gives 17-factor integrators while the second set gives 13-factor integrators as product of four S_{A_1}, six S_{A_2} and three S_{A_3}. As we can exchange A_1, A_2 and A_3, we shall take for A_2 the part for which the time-evolution mapping has the lowest cost.

3.5 Planetary Hamiltonian

Wisdom and Holman ([18]) have used a symplectic integrator for their integrations of the solar system which allowed them to use a longer step size. They used an expression of the Planetary Hamiltonian in term of Jacobi coordinates for which the Hamiltonian is splitted in two parts. Here we prefer to use the canonical heliocentric variables of Poincaré which provides a more elegant and symmetrical formulation of the hamiltonian ([12]), which is then expressed in three integrable parts : H_0, T_1, U_1.

H_0 corresponds to the sum of n disjoint Keplerian problems. T_1 is the perturbation depending only on the actions and U_1 depends only on the positions. We are thus led to search for symplectic integrators for Hamiltonians which decompose in three integrable parts $H = A + B + C$.

Let us consider O the center of mass of $n+1$ bodies of masses $m_0, \ldots, m_n$ in gravitational interaction. Let u_i be the coordinates with respect to O and and $\Delta_{i,j} = \|u_i - u_j\|$, the Hamiltonian becomes

$$H = T + U = \tfrac{1}{2} \sum_{i=0}^{n} m_i \|\dot{u}_i\|^2 - G \sum_{0 \le i < j \le n} \frac{m_i m_j}{\Delta_{i,j}}. \tag{65}$$

Let $\tilde{u}_i = m_i \dot{u}_i$, we obtain in canonical coordinates

$$T = \tfrac{1}{2} \sum_{i=0}^{n} \frac{\|\tilde{u}_i\|}{m_i}. \tag{66}$$

Let us consider now the heliocentric coordinates: $r_0 = u_0, r_i = u_i - u_0$. In order to have canonical variables, we take

$$\tilde{r}_0 = \sum_{i=0}^{n} u_i = 0, \tilde{r}_i = \tilde{u}_i, \quad 1 \le i \le n \tag{67}$$

or

$$\tilde{u}_0 = -\sum_{i=1}^{n} \tilde{r}_i, \tilde{u}_i = \tilde{r}_i, \quad 1 \le i \le n. \tag{68}$$

The kinetic energy becomes

$$T = \tfrac{1}{2} \sum_{i=1}^{n} \frac{\|\tilde{r}_i\|^2}{m_i} + \tfrac{1}{2} \frac{\|\sum_{i=1}^{n} \tilde{r}_i\|^2}{m_0} \tag{69}$$

$$= \tfrac{1}{2} \sum_{i=1}^{n} \|\tilde{r}_i\|^2 \left[\tfrac{1}{m_i} + \tfrac{1}{m_0}\right] + \sum_{0<i<j} \frac{\tilde{r}_i . \tilde{r}_j}{m_0} \tag{70}$$

and

$$U = -G \sum_{i=1}^{n} \frac{m_0 m_i}{r_i} - G \sum_{0<i<j \le n} \frac{m_i m_j}{\Delta_{i,j}}. \tag{71}$$

One can write $H = H_0 + H_1$ with $H_0 = T_0 + U_0$, $H_1 = T_1 + U_1$ where H_0 is the Hamiltonian of n disjoint two body problems: the planet of mass $\frac{m_0 m_i}{m_0 + m_i}$ around the sun of mass $m_0 + m_i$. H_1 may be considered as an interactive perturbation. We thus have

$$T_0 = \tfrac{1}{2} \sum_{i=1}^{n} \|\tilde{r}_i\|^2 \left[\tfrac{1}{m_i} + \tfrac{1}{m_0}\right], U_0 = -G \sum_{i=1}^{n} \frac{m_0 m_i}{r_i} \tag{72}$$

$$T_1 = \sum_{0<i<j} \frac{\tilde{r}_i . \tilde{r}_j}{m_0}, U_1 = -G \sum_{0<i<j \le n} \frac{m_i m_j}{\Delta_{i,j}}. \tag{73}$$

H_1 is particularly simple as the kinetic energy and the potential energy depend on coordinates and momenta respectively.

Writing $H = H_0 + T_1 + U_1$ where T_1 is very easy to integrate, we can try to use the integrator defined in (57). That is what we will do in a near future at the Bureau des Longitudes.

Published in Fields Institute Communications, **10**, 1996 19

4 Conclusion

We showed in this paper that there are exactly 5 7-factor fourth-order symplectic integrators. Three of them are known (see [7, 20]). There are exactly 46 11-factor fifth-order symplectic We showed, that there are exactly 39 15-factor reversible sixth-order symplectic integrators. All of them are reversible products of second-order integrators. Three of them were known ([20]). In the case when $h = A+B+C$, we show that minimal second-order integrator have length 5 and reversible fourth-order integrator have length 13.

Bibliography

[1] Bourbaki, N.: Groupes et algèbres de Lie, Éléments de Mathématiques, Hermann, Paris, 1972.

[2] Cary, J.R.: Lie Transform Perturbation Theory for Hamiltonian Systems, in Physics Reports, North-Holland Publishing Company **79-2** (1981), 129–159.

[3] Deprit, A.: Canonical transformations depending on a small parameter, Cel. Mech. **1** (1969), 12–30.

[4] Dragt, A. J., Finn, J. M.: Lie Series and invariant functions for analytic symplectic maps, J. Math. Physic **17** (1976), 2215–2227.

[5] Dragt, A. J., Healy, L. M.: Concatenation of Lie Algebraic Maps, in Lie Methods in Optics II, Lec. Notes in Physics **352** (1988).

[6] Finn, J. M.: Lie Series: a Perspective, Local and Global Methods of nonlinear Dynamics, Lec Notes in Physics **252** (1984), 63–86.

[7] Forest, E., Ruth, D.: Fourth-Order Symplectic Integration, Physica D **43** (1990), 105–117.

[8] P.-V. Koseleff: Calcul Formel pour les méthodes de Lie en mécanique hamiltonienne, Thèse de troisième cycle, École Polytechnique, january 1993.

[9] P.-V. Koseleff: Relations among Formal Lie Series and Construction of Symplectic Integrators, AAECC'10 proceedings, Lect. Not. Comp. Sci. **673** (1993).

[10] P.-V. Koseleff: Comparison between Deprit and Dragt-Finn perturbation methods, Celestial Mechanics **01** (1994).

[11] P.-V. Koseleff: Formal Calculus for Lie Methods in Hamiltonian Mechanics, Ph.D. Thesis, English Translation, Lawrence Berkeley Laboratory, University of California, LBID-2030, Rev. UC-405, 1994.

[12] J. Laskar: Analytical Framework in Poincaré Variables for the Motion of the Solar System, Predictability, Stability and Chaos in N-Body Dynamical Systems (1991).

[13] R. I. MacLachlan: On the numerical integration of ordinary differential equations by symmetric composition methods, SIAM J. Sci. Comp. **16(1)** (1995), 151–168.

[14] Michel, J.: Bases des Algèbres de Lie Libres, Étude des coefficients de la formule de Campbell-Hausdorff, Thèse, Orsay, 1974.
[15] Perrin, D.: Factorization of free monoids, in Lothaire M., Combinatorics On Words, Chap. 5, Addison-Wesley (1983).
[16] Steinberg, S.: Lie Series, Lie Transformations, and their Applications, in Lie Methods in Optics, Lec. Notes in Physics **250** (1985).
[17] Suzuki M.: Fractal Decomposition of Exponential Operators with Applications to Many-Body Theories and Monte Carlo Simulations, Ph. Letters A **146** (1990), 319–323.
[18] Wisdom J., Holman, M.: Symplectic Maps for the N-Body Problem, The Astr. J. **102(4)** (1991), 1528–1538.
[19] Yoshida, H.: Conserved Quantities of Symplectic Integrators for Hamiltonian Systems, Physica D (1990).
[20] Yoshida, H.: Construction Of Higher Order Symplectic Integrators, Ph. Letters A **150,** (1990), 262–268.

This article was processed using the LaTeX macro package with LLNCS style

Relations among Lie Series Transformations and Isomorphisms between free Lie Algebras

KOSELEFF, P.-V.
Discrete Maths, 180 (1998), 243-254

ELSEVIER

Discrete Mathematics 180 (1998) 243–254

DISCRETE MATHEMATICS

Relations among Lie-series transformations and isomorphisms between free Lie algebras

P.-V. Koseleff*

Équipe Analyse Algébrique, Institut de Mathématiques Université Paris 6, Case 247, 2 place Jussieu, F-75252 Paris Cedex 05, France

Received 18 July 1995; received in revised form 9 January 1997; accepted 17 February 1997

Abstract

We study the subgroup generated by the exponentials of formal Lie series. We show three different ways to represent elements of this subgroup. These elements induce Lie-series transformations. Relations among these family of transformations furnish algorithms of composition. Starting from the Lazard elimination theorem and the Witt's formula, we show isomorphisms between some submodules of free Lie algebras. Combining different results, we also show that the homogeneous terms of the Hausdorff series $H(a,b)$ freely generate the free Lie algebra $L(a,b)$ without a line.

1. Introduction

Lie-series automorphisms or Lie transformations play an important role in classical mechanics. They can be seen, for example, as the time evolution in a Hamiltonian system. The product of two such transformations may therefore be seen as the combined effects of two Hamiltonians.

The use of this formalism becomes efficient when it becomes easy to manipulate formal Lie series, to compute composition of Lie transformations or to express such transformations in several ways. They are universal identities in Lie algebras and we will work in a free Lie algebra. Instead of considering exponentials of Lie series, we will consider the group of Lie-series automorphisms. Actually after having defined the Lie transformation, historically introduced by Deprit [3], the factored product transform introduced by Dragt and Finn [4] and the exponential of an inner derivation, we will show that these transformations are the same subgroup of the Lie-series automorphisms

* E-mail: koseleff@math.jussieu.fr.

PII S0012-365X(97)00118-0

close to identity. They can be seen as conjugation in the algebra of formal Lie series. All of them are defined by generating Lie series.

After having reminded some notations in free algebras in Section 2, we will introduce formal Lie series on a weighted alphabet and define the Lie-series transformations and their properties in Section 3. In Section 4, we will consider Lie-series automorphisms they generate and their relations. In the last section, we will show several isomorphisms between free Lie algebras or subalgebras. We will prove, using combinatorial identities like the Witt's formula and a theorem of M. Lazard, that the subalgebra generated by the homogeneous terms of the Hausdorff series is isomorphic to the free Lie algebra on an alphabet of two letters without a line.

2. Notations

In this paper X will denote a weighted alphabet, that is to say an ordered set (possibly endless), in which each letter a has a non-negative integer weight $\|a\|$.

R is a ring which contains the rational numbers Q.

X^* is the free monoid generated by X. X^* is totally ordered with the lexicographic order.

$M(X)$ is the free magma generated by X.

$\mathscr{A}(X)$ is the associative algebra, that is to say the R-algebra of X^*.

$L(X)$ is the free Lie algebra on X. It is defined as the quotient of the R-algebra of $M(X)$ by the ideal generated by the elements (u,u) and $(u,(v,w))+(v,(w,u))+(w,(u,v))$. Its multiplication law $[\,,]$ is bilinear, alternate and satisfies the Jacobi identity

$$[x,[y,z]]+[y,[z,x]]+[z,[x,y]]=0. \tag{1}$$

An element of $M(X)$ considered as element of $L(X)$ will be called a Lie monomial.

By posing for $a,b\in X$, $[a,b]=ab-ba$, we have $L(X)\subset\mathscr{A}(X)$.

On $L(X)$ so as on $A(X)$, one considers the following gradations:

— Gradation by the length (the unique morphism that extends the function $a\mapsto 1$ on X). For $x\in X^*$ (resp. $M(X)$) $|x|$ denotes the length. $L_n(X)$ (resp. $\mathscr{A}_n(X)$) is the submodule generated by monomials of length n.

— One defines on X^* (resp. $M(X)$) the weight $x\mapsto\|x\|$ as the unique morphism that extends the weight on X. $\tilde{L}_n(X)$ (resp. $\tilde{\mathscr{A}}_n(X)$) is the submodule generated by monomials of weight n.

— The multi-degree is the unique morphism from X^* (resp. $M(X)$) onto $\mathbb{N}^{(X)}$ that extends $a\mapsto \mathbb{1}_a$, for $a\in X$. For a given α in $\mathbb{N}^{(X)}$, $L^\alpha(X)$ (resp. $\mathscr{A}^\alpha(X)$) denotes the submodule generated by monomials of degree α.

Remark. When $\|a\|=1$ for each $a\in X$, then obviously $L_n(X)=\tilde{L}_n(X)$ (resp. $\mathscr{A}_n(X)=\tilde{\mathscr{A}}_n(X)$).

For $x \in L(X)$, we denote by L_x the inner derivation $y \mapsto [x, y]$. The set of inner derivations of X is the adjoint Lie algebra with commutator as Lie bracket and we have from the Jacobi identity (1)

$$L_{[x,y]} = [L_x, L_y] = L_x L_y - L_y L_x. \tag{2}$$

For $x_n \in L_n(X)$, (see [2]) let $Dx_n = nx_n$. For $x_n \in \tilde{L}_n(X)$, let $\tilde{D}x_n = nx_n$. We thus define two derivations D and $\tilde{D}$ on $L(X)$. They are not inner derivations.

We define the formal Lie series $\tilde{L}(X)$ and $\tilde{\mathscr{A}}(X)$ as

$$\tilde{L}(X) = \prod_{n \geqslant 0} \tilde{L}_n(X) \quad \text{and} \quad \tilde{\mathscr{A}}(X) = \prod_{n \geqslant 0} \tilde{\mathscr{A}}_n(X).$$

We will write $x \in \tilde{L}(X)$ as a series $\sum_{n \geqslant 0} x_n$. $\tilde{L}(X)$ is a complete Lie algebra with the Lie bracket

$$([x, y])_n = \sum_{p+q=n} [x_p, y_q].$$

3. Some transformations

A transformation $T : \tilde{L}(X) \to \tilde{L}(X)$ will be called Lie-series automorphism if it is a Lie-algebra automorphism, that is to say $[Tf, Tg] = T[f, g]$. The Lie-series automorphisms act on the adjoint Lie algebra by

$$TL_f T^{-1} = L_{Tf}.$$

We give here three transformations that will give three different ways to build Lie-series automorphisms.

3.1. The exponential

Denoting by $\tilde{L}(X)^+$ (resp. $\tilde{\mathscr{A}}(X)^+$) the ideal of $\tilde{L}(X)$ (resp. $\tilde{\mathscr{A}}(X)$) generated by the elements of non-negative weight, one defines the exponential and the logarithm as

$$\exp : \tilde{\mathscr{A}}(X)^+ \to 1 + \tilde{\mathscr{A}}(X)^+,$$

$$x \mapsto \sum_{n \geqslant 0} \frac{x^n}{n!},$$

$$\log : 1 + \tilde{\mathscr{A}}(X)^+ \to \tilde{\mathscr{A}}(X)^+,$$

$$x \mapsto - \sum_{n \geqslant 1} \frac{(1-x)^n}{n}.$$

They are mutually reciprocal functions and we have (see [1, Ch. II, Section 5]) the

Theorem 1 (Campbell–Hausdorff). *For* $x, y \in \tilde{L}(X)^+$,

$$H(x,y) = \log[\exp(x)\exp(y)] \in \tilde{L}(X)^+.$$

More precisely, we have the following:

Lemma 2. *Given* $x, y \in \tilde{L}(X)^+$, *we have for* $m \geqslant 0$,

$$H(x,y)_{m+1} - x_{m+1} - y_{m+1} \in \tilde{L}_{m+1}(x_1,\dots,x_m,y_1,\dots,y_m).$$

Remark. Here $\tilde{L}_q(x_1,\dots,x_p)$ denotes the submodule generated by elements of weight q of the subalgebra generated by $\{x_1,\dots,x_p\}$.

Given $x \in \tilde{L}(X)^+$, we consider $\exp(L_x)$ defined as

$$\exp(L_x)y = \sum_{i\geqslant 0} \frac{L_x^i}{i!} y.$$

Theorem 3. *For* $x \in \tilde{L}(X)^+$, $\exp(L_x)$ *is a Lie-series automorphism* (*see* [2]). *We also have* [1,6] *for* $y \in \tilde{L}(X)$

$$\exp(x)y\exp(-x) = \exp(L_x)y,$$

$$\exp(x)\exp(y)\exp(-x) = \exp(\exp(L_x)y).$$

Proof. We have $\exp(-L_x)\exp(L_x) = \mathbb{1}$. From the Jacobi identity (1), we have ([12]) by induction on $k \geqslant 0$, for any $f, g, h \in \tilde{L}(X)^+$,

$$L_f^k[g,h] = \sum_{i=0}^{k} \binom{k}{i} [L_f^i g, L_f^{k-i} h].$$

We therefore deduce that

$$\begin{aligned}
\exp(L_f)[g,h] &= \sum_{n\geqslant 0} \frac{1}{n!} \sum_{p=0}^{n} \binom{n}{p} [L_f^p g, L_f^{n-p} h] \\
&= \sum_{p+q\geqslant 0} \frac{1}{(p+q)!} \frac{(p+q)!}{p!q!} [L_f^p g, L_f^q h] \\
&= [\exp(L_f)g, \exp(L_f)h]. \qquad \square
\end{aligned}$$

From the Theorem 1 and Eq. (2), the set $G = \{\exp(L_x), x \in L^+(X)\}$ is a group that we will call the Lie transformations group.

3.2. Factored product transform

For $g\in\tilde{L}^{+}(X)$, let us define

$$Mg=\cdots\exp(g_n)\cdots\exp(g_1).$$

Using the preceding lemmas we deduce (see [12]):

Proposition 4 (Factored product expansion). *For $k\in\tilde{L}(X)^{+}$, there is a unique $g\in\tilde{L}(X)^{+}$ such that*

$$\exp\left(\sum_{n\geqslant 1}k_n\right)=\cdots\exp(g_n)\cdots\exp(g_1)=Mg.$$

Proof. The above proposition is proved by induction, constructing $g\in\tilde{L}(X)$ and $k^{(p)}\in\prod_{n>p}\tilde{L}_n(X)$ such that, for each $p\geqslant 1$,

$$\exp(k)=\exp(k^{(p)})\exp(g_p)\cdots\exp(g_1). \qquad \square$$

Remark. This fact is also a variant of the Zassenhaus formula (see [9]).

3.3. The transformation T

We also define the transformation $T:\tilde{\mathscr{A}}(X)^{+}\to 1+\tilde{\mathscr{A}}(X)^{+}$ by

$$(Tx)_0=1,\qquad (Tx)_n=\sum_{p=1}^{n}\frac{p}{n}x_p(Tx)_{n-p},\ \ n\geqslant 1.$$

We therefore deduce that $\tilde{D}(Tx)=\tilde{D}xTx$. Conversely, the series y in $1+\tilde{\mathscr{A}}(X)^{+}$ given by

$$y_0=1,\qquad y_n=\sum_{p=1}^{n}\frac{p}{n}x_py_{n-p},\ \ n\geqslant 1,$$

is the unique solution of $\tilde{D}y=(\tilde{D}x)y$. From $(Tx)^{-1}Tx=1$, we deduce that

$$\tilde{D}((Tx)^{-1}Tx)=(\tilde{D}(Tx)^{-1})Tx+(Tx)^{-1}\tilde{D}xTx=0,$$

that is to say $\tilde{D}(Tx)^{-1}=-(Tx)^{-1}\tilde{D}x$. We thus have

$$((Tx)^{-1})_0=1,\qquad ((Tx)^{-1})_n=-\sum_{p=1}^{n}\frac{p}{n}((Tx)^{-1})_{n-p}x_p,\ \ n\geqslant 1.$$

Remark. If $[x,\tilde{D}x]=0$, then $Tx=\exp(x)$. The Lie transform appears as a generalized exponential.

3.4. Relations between transformations

Proposition 5. *Let $g \in \tilde{L}(X)^+$, there is a unique series $w \in \tilde{L}(X)^+$ such that*

$$Tw = Mg.$$

Proof. Let $x = x_n \in \tilde{L}_n(X)$, we have

$$\tilde{D}\exp(x_n) = \sum_{p \geqslant 0} \frac{1}{p!}\tilde{D}x_n^p = \sum_{p \geqslant 0} \frac{1}{p!} p n x_n^p = n x_n \exp(x_n) = n \exp(x_n) x_n.$$

We have, therefore,

$$\begin{aligned}
\tilde{D}(Mg) &= \sum_{n \geqslant 1} [\cdots \exp(g_{n+1})]\,[\tilde{D}[\exp(g_n)]][\exp(g_{n-1})\cdots\exp(g_1)] \\
&= \sum_{n \geqslant 1} [\cdots \exp(g_{n+1})]\,[n g_n \exp(g_n)]\,[\exp(g_{n-1})\cdots\exp(g_1)] \\
&= \sum_{n \geqslant 1} [\cdots \exp(g_{n+1})]\,[n g_n]\,[\exp(-g_{n+1})\cdots](Mg) \\
&= \left[\sum_{n \geqslant 1} n[\cdots \exp(L_{g_{n-1}})g_n]\right](Mg).
\end{aligned}$$

Let $\tilde{D}w = \sum_{n \geqslant 1} n[\cdots \exp(L_{g_{n+1}})]g_n$, that is to say

$$w_n = \sum_{k=1}^{n} \frac{k}{n} \sum_{\substack{(k+1)m_{k+1}+\cdots \\ +(n-k)m_{n-k}=n-k}} \frac{L_{g_{n-k}}^{m_{n-k}} \cdots L_{g_{k+1}}^{m_{k-1}}}{m_{k+1}! \cdots m_{n-k}!} g_k, \tag{3}$$

we have

$$\tilde{D}(Mg) = (\tilde{D}w)(Mg), \qquad \tilde{D}(Tw) = (\tilde{D}w)(Tw).$$

We thus deduce that $Mg = Tw$. $\square$

Eq. (3) shows that $w_n \in L(g_1, \ldots, g_n)$ and furthermore that $w_n - g_n \in L(g_1, \ldots, g_{n-1})$. These relations may be easily inverted and, combining Propositions 5 and 4, we deduce the following.

Proposition 6. *Given $w, k, g \in \tilde{L}(X)^+$, there exist*
— *$k' \in \tilde{L}(X)^+$ with $k'_n - w_n \in L(w_1, \ldots w_{n-1})$ such that $\exp(k') = Tw$,*
— *$g' \in \tilde{L}(X)^+$ with $g'_n - k_n \in L(k_1, \ldots k_{n-1})$ such that $Mg' = \exp(k)$,*
— *$w' \in \tilde{L}(X)^+$ with $w'_n - g_n \in L(g_1, \ldots g_{n-1})$ such that $Tw' = Mg$.*

4. Lie transformations

We call Lie transformation a Lie automorphism close to the identity, that is to say, which satisfies for each $a \in X$,

$$Ta - a \in \prod_{n > \|a\|} \tilde{L}_n(X).$$

Using preceding lemmas and Proposition 6, we deduce that for $x \in \tilde{L}(X)^+$,

$$\begin{aligned}
&\exp(L_x) : y \mapsto \exp(x) y \exp(-x),\\
&T_x : y \mapsto (Tx) y (Tx)^{-1},\\
&M_x : y \mapsto (Mx) y (Mx)^{-1}
\end{aligned}$$

are Lie transformations. We will show now that these three transformations are three different ways to represent the same transformation.

4.1. The Lie transform

Given $w \in \tilde{L}(X)^+$, $f \in \tilde{L}(X)$, we define $T_w f = (Tw) f (Tw)^{-1}$. We thus have

$$\begin{aligned}
\tilde{D}(T_w f) &= (\tilde{D}w)(Tw) f (Tw)^{-1} + (Tw)(\tilde{D}f)(Tw)^{-1} - (Tw) f (Tw)^{-1} \tilde{D}w\\
&= [\tilde{D}w, (Tw) f (Tw)^{-1}] + (Tw)(\tilde{D}f)(Tw)^{-1}\\
&= L_{\tilde{D}w} T_w f + T_w(\tilde{D}f).
\end{aligned} \tag{4}$$

Let $F = T_w f = \sum_{n,m \geqslant 0} F_{n,m}$, where

$$F_{n,m} = (T_w f_m)_{n+m} = \sum_{p+q=n} (Tw)_p f_m ((Tw)^{-1})_q \in \tilde{L}_{n+m}(X).$$

Using Eq. (4), we get

$$\tilde{D}F_{n,m} = (n+m) F_{n,m} = \sum_{p=1}^{n} p L_{w_p} F_{n-p,m} + m F_{n,m},$$

so

$$F_{n,m} = \sum_{p=1}^{n} \frac{p}{n} L_{w_p} F_{n-p,m}.$$

Using this algorithm, we show that $F_r = \sum_{n+m=r} F_{n,m}$ may be calculated in $O(r^2)$ Lie-brackets evaluations, by an iterative way.

4.2. Composition

Let $w_1, w_2 \in L(X)^+$ and $T = T_{w_1} T_{w_2}$. From (4) we deduce that

$$\begin{aligned}\tilde{D}(Tf) &= L_{\tilde{D}w_1} Tf + T_{w_1} \tilde{D}(T_{w_2} f)\\ &= (L_{\tilde{D}w_1} + T_{w_1} L_{\tilde{D}w_2} T_{w_1}^{-1}) Tf + T(\tilde{D}f)\\ &= L_{\tilde{D}w_1 + T_{w_1}\tilde{D}w_2} Tf + T(\tilde{D}f).\end{aligned}$$

We thus deduce that $T_{w_1} T_{w_2} = T_w$, where

$$\tilde{D}w = \tilde{D}w_1 + T_{w_1}\tilde{D}w_2. \tag{5}$$

Composition of two Lie transformations appears clearly as a Lie transformation. Furthermore, the product may be expressed as Lie transformations by an iteration algorithm, in a polynomial time of Lie-brackets evaluations. Using Lie operators or Lie-series exponentials, we should have computed the so-called Hausdorff product of w_1 and w_2.

4.3. The Dragt–Finn transform

The Dragt–Finn transform M_g is the infinite product of exponential maps (see [4]). Given $g = \sum_{n \geqslant 1} g_n$, we define M_g and M_g^{-1} as

$$M_g = \cdots \exp(L_{g_n}) \cdots \exp(L_{g_1}), \qquad M_g^{-1} = \exp(-L_{g_1}) \cdots \exp(-L_{g_n}) \cdots.$$

4.4. Relations

The three above transformations are totally defined by generating series which satisfy the following:

Proposition 7. *Given $w, k, g \in \tilde{L}(X)^+$, the series defined in Proposition 6 satisfy*

$$\exp(L_{k'}) = T_w, \qquad T_{w'} = M_g, \qquad M_{g'} = \exp(L_k).$$

Remark. We deduce in passing that the Lie transform is a Lie-series automorphism close to identity and that any Lie transformation $T \in G$ may be expressed as an exponential of a Lie operator or as an infinite product of single exponentials or as a proper Lie transform. The use of a representation depends deeply on the result we look for. For example, if we have to compose transformations, it is much easier to consider Lie transforms because their product is a Lie transform whose generating function appears easily from (5).

We will not explain in this paper how to compute explicitly the relations between these transformations, but that can be made, using the Lyndon basis and does not require to go through the associative algebra (see [6]).

Proposition 7 may be turned into:

Proposition 8. *Given* $w, g, k \in \tilde{L}(X)^+$, *such that*

$$\exp(L_k) = T_w = M_g,$$

then for each $n \in \mathbb{N}$, *we have*

$$w_n - k_n \in \tilde{L}_{n-1}(X), \qquad w_n - g_n \in \tilde{L}_{n-1}(X), \qquad g_n - k_n \in \tilde{L}_{n-1}(X),$$

and

$$\tilde{L}(w_1, \ldots, w_n) = \tilde{L}(k_1, \ldots, k_n) = \tilde{L}(g_1, \ldots, g_n).$$

5. Free Lie-algebras isomorphisms

Let us first remind the elimination theorem of Lazard [1].

Theorem 9. *Let* $S \subset X$ *and*

$$T = \{(s_1, \ldots, s_n, x),\ n \geqslant 0,\ s_1, \ldots, s_n \in S,\ x \in X - S\}.$$

— $L(X)$ *is the direct sum of* $L(X - S)$ *and of the ideal* $\mathscr{S}$ *generated by* S.
— $L(T)$ *and* $\mathscr{S}$ *are isomorphic through* $(s_1, \ldots, s_n, x) \mapsto L_{s_1} \cdots L_{s_n} x$.

By considering $X = \{a, b\}$ and $S = \{a\}$, we get the following isomorphism

$$L(\{a,b\}) = L(\{a\}) \oplus L(\{L_a^n b,\ n \geqslant 0\}) = K\,.\,a \oplus L(\{L_a^n b,\ n \geqslant 0\}). \tag{6}$$

By posing $X = \{L_a^k b,\ k \geqslant 0\}$ and $S = \{L_a^k b,\ k \geqslant p\}$, we deduce that

$$\begin{aligned} L(\{a,b\}) &= K\,.\,a \oplus L(\{L_a^k b,\ k \geqslant 0\}) \\ &= K\,.\,a \oplus L(\{L_a^k b,\ 0 \leqslant k \leqslant p-1\}) \oplus (L_a^k b,\ k \geqslant p). \end{aligned}$$

We, therefore, conclude that

$$\begin{aligned} L(\{a,b\})/(L_a^p b) &= L(\{a,b\})/(L_a^k b,\ k \geqslant p) \\ &= K\,.\,a \oplus L(\{L_a^k b,\ 0 \leqslant k \leqslant p-1\}). \end{aligned}$$

That proves that the algebra generated by $\{a, b,\ L_a^p b = 0\}$ is isomorphic to the weighted free Lie algebra $L(\{L_a^k b,\ 0 \leqslant k \leqslant p-1\})$ and the line generated by a.

We will show now that these isomorphisms are isomorphisms between homogeneous submodules.

5.1. Dimension of the homogeneous components

Let us first remind some well-known identities. Given an indexed alphabet X, we consider the dimension $l(\alpha)$ of $L^\alpha(X)$. Using the following identity between formal series (see [1]), which results from the Poincaré–Birkhoff–Witt's theorem

$$1 - \sum_{x\in X} T_x = \prod_{\alpha\in\mathbb{N}^{(X)}-\{0\}} (1 - T^\alpha)^{l(\alpha)}, \tag{7}$$

we deduce that

$$l(\alpha) = \frac{1}{|\alpha|}\sum_{d|\alpha}\mu(d)\frac{(|\alpha/d|)!}{((\alpha/d)!)} \quad \text{or} \quad \sum_{\beta|\alpha}|\beta|\,l(\beta) = \frac{|\alpha|!}{\alpha!}.$$

Let us take now the gradation by the length and calculate $l_n = \sum_{|\alpha|=n} l(\alpha)$, the dimension of $L_n(X)$. As [1], let us substitute in (7) the same unknown U to T_x, we get for a finite alphabet of cardinality q:

$$1 - qU = \prod_{\alpha\in\mathbb{N}^X-\{0\}} (1 - U^{|\alpha|})^{l(\alpha)} = \prod_{r>0}(1 - U^r)^{l_r} \tag{8}$$

that is to say, the Witt's formula [1]: $\sum_{d|n} d\,l_d = q^n$.

Let $X = \biguplus_{p\geqslant 1} X_p$ be a weighted alphabet where each letter of X_p has a weight p. Let $\tilde{l}_n = \sum_{\|\alpha\|=n} l(\alpha)$ be the dimension of $\tilde{L}_n(X)$. If X_i has a cardinality q_i, let us substitute in (7) U^i to T_x for $x\in X_i$. We thus obtain

$$1 - \sum_{i\geqslant 1} q_i U^i = \prod_{r>0}\prod_{\|\alpha\|=r}(1 - U^{\|\alpha\|})^{l(\alpha)} = \prod_{r>0}(1 - U^r)^{\tilde{l}_r}.$$

In the particular case where $q_i = p$ for each $i\in\mathbb{N}$, we thus deduce

$$\prod_{r>0}(1 - U^r)^{\tilde{l}_r} = 1 - p\sum_{i>0} U^i = \frac{1-(p+1)U}{1-U}. \tag{9}$$

From identities (8) and (9), we then obtain

Isomorphism 1. *Let* $X = \{x_1,\ldots,x_q\}$ *and* $Y = \biguplus_{i\geqslant 1} Y_i$, *where* $\operatorname{Card} Y_i = q-1$. *We have*

$$\dim L_1(X) = q, \qquad \dim \tilde{L}_1(Y) = q-1, \qquad \dim L_n(X) = \dim \tilde{L}_n(Y), \quad n\geqslant 2,$$

that can be also expressed as

$$\sum_{d|n} d\dim L_d(X) = q^n, \qquad \sum_{d|n} d\dim \tilde{L}_d(Y) = q^n - 1.$$

In the particular case where $q=2$, we recover the isomorphism defined in Theorem 9, by posing $Y = \{L_a^p b,\ p\geqslant 0\}$.

5.2. The Hausdorff series

Let us suppose now that $X = \{a,b\}$ and $\|a\| = \|b\| = 1$. Let $H(a,b)$ the Hausdorff series of a,b defined in Theorem 1. We have

$$\exp(H(a,b)) = \exp(a)\exp(b).$$

From the definition in Section 3.3 and the remark in the proof of Proposition 5, we have $Ta = \exp(a)$ and $Tb = \exp(b)$. Let $G(a,b)$ be the solution of $TG(a,b) = \exp(a)\exp(b)$, we get using relation (5)

$$\tilde{D}G(a,b) = DG(a,b) = Da + T_a Db = Da + \exp(L_a)Db,$$

that is to say

$$G(a,b) = a + b + \sum_{n\geqslant 1} \frac{1}{(n+1)!} L_a^n b.$$

We can now prove the following result:

Isomorphism 2. *Let $X = \{a,b\}$ and $H(a,b) = \sum_{n\geqslant 1} H_n$. The subalgebra $L(\{H_n,\ n\geqslant 0\})$ is isomorphic to the free Lie algebra $L(\{L_b^n a,\ n\geqslant 0\})$.*

Proof. Using Proposition 8, we know that for $d\geqslant 1$,

$$\tilde{L}_d(\{G_n(a,b),\ n\geqslant 1\}) = \tilde{L}_d(\{H_n(a,b),\ n\geqslant 1\}).$$

But $G_n(a,b) = \frac{1}{n!} L_a^{n-1} b$, and from Isomorphism 1, we know that the subalgebra $\tilde{L}(\{L_a^n b,\ n\geqslant 0\})$ is free and that

$$\tilde{L}_d(\{L_a^n b,\ n\geqslant 0\}) = L_d(\{a,b\}), \quad d\geqslant 2.$$

We thus deduce that the subalgebra generated by the homogeneous terms of the Hausdorff series is free and therefore isomorphic to the free Lie algebra $L(\{a,b\})$ without a line. $\square$

Remark. Since Sirsov and Witt (see [11, Theorem 2.5]), it is known that $\tilde{L}_d(\{H_n(a,b),\ n\geqslant 1\})$ is free. Here we proved that $\{H_n(a,b)\}$ freely generate $L(\{L_a^n b,\ n\geqslant 0\})$.

6. Conclusions

We have shown in this in paper how to express any transformation that belongs to the subgroup of Lie transformations in three different ways. In Hamiltonian mechanics this subgroup is exactly the group of Lie-series automorphisms close to identity. These methods have many applications like the search of the so-called symplectic integrators that are numerical methods to integrate dynamical systems [7, 13]. Using this formalism, one can also compute formal first integral for perturbed hamiltonian systems [3, 6, 12]. Regards to the computational cost, these methods have the advantage that all the series we manipulate are formal Lie series. It avoids calculations in the associative algebra [14] and the use of the Poincaré–Birkhoff–Witt basis [10].

References

[1] N. Bourbaki, Groupes et algèbres de Lie Éléments de Mathématiques Hermann, Paris, 1972.

[2] P. Cartier, Démonstration algébrique de la formule de Hausdorff, Bull. Soc. Math. France 84 (1956) 241–249.

[3] A. Deprit, Canonical transformations depending on a small parameter, Celestial Mech. 1 (1969) 12–30.

[4] A.J. Dragt, J.M. Finn, Lie Series and invariant functions for analytic symplectic maps, J. Math. Phys. 17 (1976) 2215–2227.

[5] J.M. Finn, Lie Series: a Perspective, Local and Global Methods of nonlinear Dynamics, Lecture Notes in Physics, vol. 252, Springer, Berlin, 1984, pp. 63–86.

[6] P.-V. Koseleff, Calcul Formel pour les méthodes de Lie en mécanique hamiltonienne, Thèse, École Polytechnique, 1993.

[7] P.-V. Koseleff, Relations among Formal Lie Series and Construction of Symplectic Integrators, AAECC'10 Proc., Lecture Notes in Computer Science, vol. 673, Springer, Berlin, 1993.

[8] P.-V. Koseleff, Comparison between Deprit and Dragt–Finn perturbation methods, Celestial Mech. 58 (1) (1994).

[9] Magnus et al., Combinatorial Group Theory: Presentation of Groups in Terms of Generators and Relations, Wiley, New York, 1966.

[10] M. Petitot, Algèbre non commutative en Scratchpad: application au problème de la réalisation minimale analytique, Thèse, Université de Lille I, 1991.

[11] Reutenauer, Ch., Free Lie Algebras, Oxford University Press, Oxford, 1993.

[12] S. Steinberg, Lie Series, Lie Transformations, and their Applications, in: Lie Methods in Optics, Lecture Notes in Physics, vol. 250, Springer, Berlin, 1985.

[13] M. Suzuki, General Theory of higher-order decomposition of exponential operators and symplectic integrators, Phys. Lett. A 165 (1992) 387–395.

[14] G. Viennot, Algèbres de Lie libres et Monoïdes Libres, Lecture Notes in Mathematics, vol. 691, Springer, Berlin, 1978.

Elementary Approximation of Exponential of Lie Polynomials

F. JEAN, P.-V KOSELEFF

AAECC'12 Proceedings, Lecture Notes in Computer Science **1255**, 1997, pp. 174-188, Springer

Elementary Approximation of Exponentials of Lie Polynomials ⋆

Frédéric JEAN & Pierre–Vincent KOSELEFF

Équipe Analyse Algébrique, Institut de Mathématiques
Université Paris 6, Case 82
4 place Jussieu, F-75252 Paris Cedex 05

e-mail : [jean,koseleff]@math.jussieu.fr

Abstract. Let $\mathcal{L} = L(x_1, \ldots, x_m)$ be a graded Lie algebra generated by $\{x_1, \ldots, x_m\}$. In this paper, we show that for any element P in $\mathcal{L}$ and any order k, $\exp(P)$ may be approximated at the order k by a finite product of elementary factors $\exp(\lambda_i x_i)$. We give an explicit construction that avoids any calculation in the Lie algebra.

0 Introduction

In hamiltonian mechanics, the phase space is governed by an Hamiltonian h and the equations $\dot{z}_i = \{z_i, h\}$ where $\{,\}$ is the Poisson bracket. The set of smooth functions on the phase space is turned into Lie algebra by considering the Poisson bracket. Integrations of hamiltonian flows by numerical schemes make use of the so-called symplectic integrators that preserve some invariants (see [8, 12, 13]). One will try to approximate the flow $\exp(t\{\cdot, A+B\})$ by composition of the hamiltonian flows of $\exp(t\{\cdot, A\})$ of A and $\exp(t\{\cdot, B\})$ of B. These methods are used for their stability in very long-time integration problems. Such integrators may be found by considering universal identities in free Lie algebras (see [4, 5, 13, 14]).

In control theory, for a control system $\dot{x} = \sum_{i=1}^{m} u_i(t) X_i(x)$, the classical problem of motion planning is the following (see [3, 6, 7]): *given two states p and q, find a feasible trajectory (i.e. the controls $u_1(t), \ldots, u_m(t)$) that steers the system from p to a point arbitrarily close to q.*

Let us assume that q is given as $\exp(X)p$, where X belongs to the Lie algebra generated by the vector fields X_i. We are interested in the simplest trajectories, those obtained as composition of the flows of the X_i's. The end point of such a trajectory is written $\exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s})p$, that is a product of elementary factors applied to the state p. A solution of our problem can then be obtained by approximating the exponential $\exp(X)$ by a product of elementary factors $\exp(\lambda X_i)$.

In this paper we will see such approximations as universal approximations in graded Lie algebra. That is why we will work in free Lie algebras and look for

⋆ Proceedings of AAECC'12 Conference, Toulouse, juin 1997, LNCS **1255**, 15 p.

universal identities in free Lie algebras. In Section **1**, we will first introduce some notations and set our result in Part 1.3. In Section **2** we will describe several independent steps of the algorithm and finally we will discuss in Section **3** the accuracy of our algorithm by considering some examples and giving some bounds for the complexity. We will finally study the application of our result to the motion planning problem.

1 Notations and Definitions

In this section we recall some basic results about Lie algebra. To keep the paper easy to read, we will not go too deeply into the theory and will not present the results in a classical way. For instance we present Lemma 3 as a consequence of the Campbell-Hausdorff Theorem and we will not tell about central integer filtrations (although our result is based only on the properties of these filtrations). For a more classical presentation of this theory we refer to [1, Ch. II] and [10].

1.1 Notations

In this paper A will denote an ordered alphabet (possibly endless).

A^* is the free monoid generated by A (the set of words). A^* is totally ordered with the lexicographic order.

$M(A)$ is the free magma generated by A (the set of parenthesed words). Having defined $M_1(A)$ as A, we define $M_n(A)$ by induction on n:

$$M_n(A) = \bigcup_{p+q=n} M_p(A) \times M_q(A) \quad \text{and} \quad M(A) = \bigcup_{n \geq 1} M_n(A).$$

$\mathcal{A}(A)$ is the associative algebra, that is to say the $\mathbb{R}$-algebra of A^*.

$L(A)$ is the free Lie algebra on A. It is defined as the quotient of the $\mathbb{R}$-algebra of $M(A)$ by the ideal generated by the elements (u, u) and $(u, (v, w)) + (v, (w, u)) + (w, (u, v))$. Its multiplication law $[,]$ is bilinear, alternate and satisfies the Jacobi identity

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0.$$

An element of $M(A)$ considered as element of $L(A)$ will be called a Lie monomial. Elements of $L(A)$ will be called Lie polynomials.

By setting $[x, y] = xy - yx$ for $x, y \in A$, we have $L(A) \subset \mathcal{A}(A)$. We will also denote by ad x, the map $y \mapsto [x, y]$.

$\mathcal{A}(A)$, and hence $L(A)$, are graded by the length (the unique morphism that extends the function $x \mapsto 1$ on A). For $x \in A^*$ (resp. $M(A)$) $|x|$ denotes the length. $L_n(A)$ (resp. $\mathcal{A}_n(A)$) is the submodule generated by monomials of length n.

We define $\hat{L}(A)$ and $\hat{\mathcal{A}}(A)$ as

$$\hat{L}(A) = \prod_{n\geq 1} L_n(A) \quad \text{and} \quad \hat{\mathcal{A}}(A) = \prod_{n\geq 1} \mathcal{A}_n(A).$$

We will write $x \in \hat{L}(A)$ as a series $\sum_{n\geq 0} x_n$. $\hat{L}(A)$ is a complete Lie algebra with the Lie bracket

$$([x,y])_n = \sum_{p+q=n} [x_p, y_q].$$

We will also use

$$\hat{L}_{\geq p}(A) = \prod_{n\geq p} L_n(A) \quad \text{and} \quad \hat{\mathcal{A}}_{\geq p}(A) = \prod_{n\geq p} \mathcal{A}_n(A).$$

1.2 Exponentials

One defines the exponential and the logarithm as

$$\begin{array}{rlrl} \exp : \hat{\mathcal{A}}(A) & \to 1 + \hat{\mathcal{A}}(A) & \log : 1 + \hat{\mathcal{A}}(A) & \to \hat{\mathcal{A}}(A) \\ x & \mapsto \sum_{n\geq 0} \dfrac{x^n}{n!}, & x & \mapsto -\sum_{n\geq 1} \dfrac{(1-x)^n}{n}. \end{array}$$

They are mutually reciprocal functions and we have (see [1, Ch. II, §5]):

Theorem 1 (Campbell-Hausdorff). *For $a, b \in A$, let $H(a,b)$ such that:*

$$\exp(a)\exp(b) = \exp(H(a,b))$$

Then $H(a,b) \in \hat{L}_{\geq 1}(A)$ and $H_1(a,b) = a+b$, $H_2(a,b) = \frac{1}{2}[a,b]$.

Therefore, for $x, y \in \hat{L}$, we have

$$\exp(H(x,y)) = \exp(x)\exp(y).$$

and $(H(x,y))_k - (x_k + y_k) \in L(x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1})$, (the free subalgebra of $\hat{L}(A)$ generated by $\{x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1}\}$)

Remark. — $H(x,y)$ is the series $H(a,b)$ evaluated at $x = a$ and $y = b$. The Hausdorff series $H(x,y)$ may be explicitly computed (see [4]).

1.3 Setting of the Main Result

Let us set $\hat{F} = \exp(\hat{L})$ and $\hat{F}_{\geq p}(A) = \exp(\hat{L}_{\geq p}(A))$. As exp is bijective, we deduce from the Campbell-Hausdorff Theorem that the sets $\hat{F}$ and $\hat{F}_{\geq p}(A)$ are some groups.

As

$$[\exp(\lambda_1 a_1) \cdots \exp(\lambda_n a_n)]^{-1} = \exp(-\lambda_n a_n) \cdots \exp(-\lambda_1 a_1),$$

the set G generated by the family $\{\exp(\lambda a), \lambda \in \mathbb{R}, a \in A\}$ is a subgroup of $\hat{F}$.

We will show that the elements of $\exp(L(A))$ can be approximated by elements of G. Let us first clarify the meaning of approximation. If $P \in L(A)$ is a Lie polynomial, we say that $\alpha = \varphi_k(P)$ is a kth-order approximation of $\exp(P)$ if:

$$\begin{aligned} \alpha = \exp(\lambda_1 a_1) \cdots \exp(\lambda_m a_m) &\in G, \\ \exp(-P)\alpha &\in \hat{F}_{\geq k+1}(A). \end{aligned} \tag{1}$$

Using Theorem 1, we thus have

$$\varphi_k(P) = \exp(P)\exp(R_{\geq k+1}) = \exp(P + R'_{\geq k+1})$$

where $R_{\geq k+1}$ and $R'_{\geq k+1} \in \hat{L}_{\geq k+1}(A)$. Both characterizations will be used in the sequel.

The aim of this paper is to prove the following theorem (Theorem 8):

Theorem. *Let P be a Lie polynomial of degree n. Then there exists an approximation of $\exp(P)$ at any order $k \geq n$.*

Furthermore we will give here an explicit construction of an approximation $\varphi_k(P)$. We will use the following steps

1. Define an approximation $\varphi_n(\lambda m)$ for any $\lambda \in \mathbb{R}$ and any Lie monomial $m \in M_n(A)$. This approximation will have the order $n = |m|$.
2. For any homogeneous Lie polynomial P of degree n, define an approximation $\varphi_n(P)$ of order n.
3. For any homogeneous Lie polynomial, show how to build a kth-order approximation $\varphi_k(P)$ from the approximation $\varphi_n(P)$. This process will depend only on n and k.
4. For any polynomial P, show how to get a kth-order approximation. This makes use of the factored product expansion

   $$\exp(P) = \exp(P_1) \cdots \exp(P_k)\exp(R_{\geq k+1}),$$

   where $P_i \in L_i(A)$ and $R_{k+1} \in \hat{L}_{\geq k+1}(A)$. This factorization is not given explicitly but may be computed using the Hausdorff series that may also be computed.

2 Approximations of Exponential of Lie Polynomials

Before going more deeply into the method of approximation, we need some results on the groups $\hat{F}$ and G.

2.1 The Groups $\hat{F}$ and G

Let $\alpha \in G$. As element of $\hat{F}_{\geq 1}(A)$, α is written as $\exp(\sum_{n\geq 1} x_n)$, with $x_n \in L_n(A)$, and the inverse of α is:

$$\alpha^{-1} = \exp(-\textstyle\sum_{n\geq 1} x_n).$$

Given $x \in \hat{L}$, we consider $\exp(\text{ad } x)$ defined as

$$\exp(\text{ad } x)y = \sum_{i\geq 0} \frac{(\text{ad } x)^i}{i!} y. \tag{2}$$

Lemma 2. *For $x \in \hat{L}$, we have (see [1])*

$$\exp(x)\exp(y)\exp(-x) = \exp(\exp(\text{ad } x)y).$$

We thus deduce the well-known lemma

Lemma 3. *Given $\alpha = \exp(x) \in \hat{F}_{\geq p}(A)$, $\beta = \exp(y) \in \hat{F}_{\geq q}(A)$, we have*

$$\alpha\beta\alpha^{-1}\beta^{-1} = \exp(z) \in \hat{F}_{\geq p+q}(A) \quad \textit{and} \quad z_{p+q} = [x_p, y_q].$$

Proof. — Let $\alpha = \exp(\sum_{k\geq p} x_k)$, $\beta = \exp(\sum_{k\geq q} y_k)$. We have

$$\begin{aligned}\alpha\beta\alpha^{-1}\beta^{-1} &= \exp(x)\exp(y)\exp(-x)\exp(-y)\\ &= \exp(\exp(\text{ad } x)y)\exp(-y) = \exp(H(\exp(\text{ad } x)y, -y)).\end{aligned}$$

But

$$\begin{aligned}H(\exp(\text{ad } x)y, -y) &= \exp(\text{ad } x)y - y + \sum_{k\geq 2} H_k(\exp(\text{ad } x)y, -y))\\ &= [x, y] + \sum_{n\geq 2} \frac{(\text{ad } x)^n}{n!} y + \sum_{k\geq 2} H_k(\exp(\text{ad } x)y, -y)).\end{aligned}$$

For each $n \geq 2$, $(\text{ad } x)^n y \in \hat{L}_{\geq np+q}(A) \subset \hat{L}_{\geq 2p+q}(A)$.

For each $k \geq 2$, $H_k(\exp(\text{ad } x)y, -y) = H_k(\exp(\text{ad } x)y - y, -y)$. But $\exp(\text{ad } x)y - y \in \hat{L}_{\geq p+q}(A)$, so $H_k(\exp(\text{ad } x)y - y, -y) \in \hat{L}_{\geq p+2q}(A)$. □

For $t \in \mathbb{R}$, we define the morphism of algebra

$$\begin{aligned} \phi_t : \hat{\mathcal{A}}(A) &\to \hat{\mathcal{A}}(A) \\ \sum_{n\geq 1} x_n &\mapsto \sum_{n\geq 1} t^n x_n . \end{aligned} \tag{3}$$

If $\alpha = \exp(\lambda_1 a_1)\cdots\exp(\lambda_n a_n) \in G \subset 1 + \hat{\mathcal{A}}(A)$, then we denote by $\alpha[t]$

$$\alpha[t] = \phi_t(\alpha) = \exp(t\lambda_1 a_1)\cdots\exp(t\lambda_n a_n) \in G. \tag{4}$$

2.2 Approximation at the Order n

Lemma 4. *Let x be a Lie monomial of length n. Then for each λ in $\mathbb{R}$ there exists an approximation of* $\exp(\lambda x)$ *at order n.*

Proof. — The proof and also the construction of the approximation is given by induction on $|x|$.

If $x \in A$, $\exp(\lambda x)$ is obviously an approximation of itself at any order. Let us assume that the result is true for the Lie monomials of length $\leq n$ and let x be a monomial of length $n+1$. In a canonical way we can write x as $[a, b]$ with a and b monomials of length p and $q \leq n$ such that $p + q = n + 1$. We have also $\lambda x = [\lambda a, b]$ for any $\lambda \in \mathbb{R}$.

By induction hypothesis there exist $\alpha, \beta \in G$ such that:

$$\begin{aligned} \alpha &= \exp(\lambda a + R_{\geq p+1}),\ R_{\geq p+1} \in \hat{L}_{\geq p+1}(A), \\ \beta &= \exp(b + R_{\geq q+1}),\ R_{\geq q+1} \in \hat{L}_{\geq q+1}(A). \end{aligned}$$

We deduce from Lemma 3 that

$$\gamma = \alpha\beta\alpha^{-1}\beta^{-1} = \exp(\lambda[a, b] + R),\ R \in \hat{L}_{\geq p+q+1}(A) = \hat{L}_{\geq n+2}(A).$$

Thus γ is an approximation of $\exp(\lambda x)$ at order $n + 1$ and the induction is done. ∎

Lemma 5. *Let P and Q be homogeneous Lie polynomials of degree $n \geq 1$ and assume that $\varphi_n(P)$ and $\varphi_n(Q)$ are approximations of* $\exp(P)$ *and* $\exp(Q)$ *at order n. Then*

$$\varphi_n(P + Q) = \varphi_n(P)\varphi_n(Q)$$

is an approximation of $\exp(P + Q)$ *at order n.*

Proof. — The approximations $\varphi_n(P)$ and $\varphi_n(Q)$ can be written as

$$\varphi_n(P) = \exp(P + R),\ \varphi_n(Q) = \exp(Q + R'),\ R, R' \in \hat{L}_{\geq n+1}(A).$$

Using the Campbell-Hausdorff Theorem, we get $\varphi_n(P)\varphi_n(Q) = \exp(H(P + R, Q + R'))$, where

$$H(P + R, Q + R') = P + Q + (R + R') + \sum_{k\geq 2} H_k(P + R, Q + R'). \qquad (5)$$

But $R + R' \in \hat{L}_{\geq n+1}(A)$ and $H_k(P + R, Q + R') \in \hat{L}_{\geq kn}(A) \subset \hat{L}_{\geq n+1}(A)$ when $k \geq 2$ so there is $R'' \in \hat{L}_{\geq n+1}(A)$ such that

$$\varphi_n(P)\varphi_n(Q) = \exp(P + Q + R'')$$

is a nth-order approximation of $\exp(P + Q)$. $\square$

Corollary 6. *If P is an homogeneous Lie polynomial of degree $n \geq 1$, then there exists an approximation of* $\exp(P)$ *at order n.*

Proof. — Let $P \in L_n(A)$. One can write $P = \sum_{i=1}^{d} \lambda_i m_i$, where the m_i's are Lie monomials and by induction we get, using Lemma 5

$$\varphi_n(P) = \varphi_n(\lambda_1 m_1) \cdots \varphi_n(\lambda_d m_d). \qquad \square$$

Note that the m_i's are not unique. That proves only the existence of nth-order approximants.

2.3 Approximations of Homogeneous Lie Polynomials

Lemma 7. *Let P be an homogeneous Lie polynomial of degree n. Then, for any $k \geq n$, there exists a kth-order approximation of* $\exp(P)$.

Proof. — For a given polynomial P of degree n, we proceed by induction on k, following Suzuki's idea ([12]). The case $k = n$ has already been done in Corollary 6. Let us assume that, for some $k \geq n+1$, there exists a $(k-1)$th-order approximation $\alpha = \varphi_{k-1}(P)$, that is $\alpha = \exp(P + \sum_{i\geq k} R_i)$ with $R_i \in L_i(A)$.

We have to distinguish two cases.

- If k is odd, we set $\gamma = \alpha[t]\alpha[s]\alpha[t]$ and we will show that for some t and s, γ is a kth-order approximation of $\exp(P)$. From Formula (5) and applying Theorem 1 twice we get

  $$\gamma = \exp((2t^n + s^n)P + (2t^k + s^k)R_k + \textstyle\sum_{i\geq k+1} R'_i).$$

  By setting

  $$t = (2 + (-1)^n 2^{n/k})^{-1/n},\ s = -2^{1/k} t,$$

  we have $2t^n + s^n = 1$ and $2t^k + s^k = 0$ (notice that t is defined since $k > n$). With these values, $\gamma = \varphi_k(P)$ is an approximation of $\exp(P)$ at order k.

- If k is even, we set $\gamma = \alpha[u]\alpha[v]^{-1}\alpha[u]$. In the same way as for the odd case, we have

$$\gamma = \exp((2u^n - v^n)P + (2u^k - v^k)R_k + \textstyle\sum_{i\geq k+1} R_i''),$$

and we obtain $2u^n - v^n = 1$ and $2u^k - v^k = 0$ by setting

$$u = (2 - 2^{n/k})^{-1/n}, \; v = 2^{1/k}u. \qquad \square$$

2.4 Approximation of any Lie Polynomial

Theorem 8. *Let P be a Lie polynomial of degree n. Then there exists an approximation of* $\exp(P)$ *at any order $k \geq n$.*

Proof. — This theorem will be proved by using the following result ([11])

Lemma 9. *Let $P \in L(A)$. Then for all $k \geq 1$ there exists $P_1 \in L_1(A), \ldots, P_k \in L_k(A)$ and a remainder $R_{\geq k+1} \in \hat{L}_{\geq k+1}(A)$ such that*

$$\exp(P) = \exp(P_1) \cdots \exp(P_k) \exp(R_{\geq k+1}). \tag{6}$$

Remark. — This result is also a variant of the Zassenhaus formula (see [9]).

The factorization (6) is constructed by induction on k. If $R_{\geq k+1} = R_{k+1} + R_{k+2} + \cdots$ is the $(k+1)$th-order remainder, the $(k+2)$th-order remainder $R_{\geq k+2}$ is given by

$$\exp(R_{\geq k+2}) = \exp(-R_{k+1}) \exp(R_{\geq k+1}).$$

Theorem 1 ensures that $R_{\geq k+2}$ belongs to $\hat{L}_{\geq k+2}(A)$ and, if we set $P_{k+1} = R_{k+1}$, the induction is done.

Let us fix now $k \geq n$ and consider the factorization (6) of P. Each term P_i is an homogeneous Lie polynomial of degree $i \leq k$. Lemma 7 can then be applied and we set

$$\begin{aligned}\alpha &= \varphi_k(P_1) \cdots \varphi_k(P_k) \\ &= \exp(P_1) \exp(R^1_{\geq k+1}) \cdots \exp(P_k) \exp(R^k_{\geq k+1}).\end{aligned}$$

From Lemma 2, we have, if $R \in \hat{L}_{\geq k+1}(A)$:

$$\begin{aligned}\exp(-P_i) \exp(R) \exp(P_i) &= \exp(\exp(\mathrm{ad}\,(-P_i))R) \\ &= \exp(R')\end{aligned}$$

where $R' \in \hat{L}_{\geq k+1}(A)$ (see Formula (2)). That means that $\exp(R)\exp(P_i) = \exp(P_i)\exp(R')$ and so, using this identity k times in the expression of α, we obtain:

$$\begin{aligned}\alpha &= \exp(P_1) \cdots \exp(P_k) \exp(R') \\ &= \exp(P) \exp(-R_{\geq k+1}) \exp(R'),\end{aligned}$$

where $R' \in \hat{L}_{\geq k+1}(A)$. Thus α is an approximation of $\exp(P)$ at order k. $\square$

2.5 Algorithm

The proof of Theorem 8 is a constructive proof. It allows the construction of an effective approximation of a Lie polynomial. Let us detail the construction of $\varphi_k(P)$, kth-order approximation of $\exp(P)$. As an input, we shall give $P_1, \ldots, P_k$ such that

$$\exp(P) = \exp(P_1) \cdots \exp(P_k) \exp(R_{\geq k+1}).$$

Each P_i is given as an explicit combination of Lie monomials. As output we will have a list $\{(\lambda_i, a_i)\}$ such that

$$\varphi_k(P) = \exp(\lambda_1 a_1) \cdots \exp(\lambda_m a_m).$$

1. For $a \in A, \lambda \in \mathbb{R}$ and $k \geq 1$, we set $\varphi_k(\lambda a) = \exp(\lambda a)$.
2. For a monomial x of length n, we write x as a bracket $[a, b]$ with $|a| = p$ and $|b| = n - p$ (this decomposition need not be unique). We thus define $\varphi_n(\lambda x)$ by the induction formula (see Lemma 4):
$$\varphi_n(\lambda x) = \varphi_p(\lambda a)\varphi_{n-p}(b)\varphi_p(\lambda a)^{-1}\varphi_{n-p}(b)^{-1}.$$
3. For an homogeneous polynomial $P = \sum_{i=1}^{d} \lambda_i x_i \in L_n(A)$, we set (see Corollary 6):
$$\varphi_n(P) = \varphi_n(\lambda_1 x_1) \cdots \varphi_n(\lambda_d x_d).$$
4. For an homogeneous polynomial $P \in L_n(A)$ and $k \geq n+1$, the approximation is defined by the recursion formulae (see Lemma 7):
$$\text{if } k \text{ is odd, then } \varphi_k(P) = \varphi_{k-1}(P)[t]\ \varphi_{k-1}(P)[-2^{1/k}t]\ \varphi_{k-1}(P)[t],$$
$$\text{if } k \text{ is even, then } \varphi_k(P) = \varphi_{k-1}(P)[u]\ \varphi_{k-1}(P)[2^{1/k}u]^{-1}\ \varphi_{k-1}(P)[t]$$
where $t = (2 + (-1)^n 2^{n/k})^{-1/n}$ and $u = (2 - 2^{n/k})^{-1/n}$.
5. For a polynomial P given by $P_1, \ldots, P_k$ such that
$$\exp(P) = \exp(P_1) \cdots \exp(P_k) \exp(R_{\geq k+1}),$$
we set (see Theorem 8):
$$\varphi_k(P) = \varphi_k(P_1) \cdots \varphi_k(P_k).$$

2.6 A Short Example

Let $P = a + [b, a]$. We have $\exp(P) = \exp(a)\exp([b, a])\exp(R_{\geq 3})$. Then we find

$$\varphi_2(a) = \exp(a), \varphi_2([b, a]) = \varphi_2(b)\varphi_2(a)\varphi_2(b)^{-1}\varphi_2(a)^{-1}$$

and thus

$$\varphi_2(P) = \exp(a)\exp(b)\exp(a)\exp(-b)\exp(-a).$$

For the same polynomial $P = a - [a, b]$, we found a shorter solution

$$\varphi_2(P) = \exp(b)\exp(a)\exp(-b).$$

3 Estimations and Example

In this section we will discuss the shape of the output of our algorithm. We will first estimate the number of elementary factors in the approximation (an elementary factor is a term $\exp(\lambda a)$, with $\lambda \in \mathbb{R}$, $a \in A$). Then we will show how one can reduce this complexity.

3.1 Complexity

Let P be a Lie polynomial of degree n and an integer $k \geq n$. We assume that we know the decomposition (6)

$$\exp(P) = \exp(P_1) \cdots \exp(P_k) \exp(R_{\geq k+1}),$$

where each polynomial $P_i = \sum_{j=1}^{d_i} \lambda_j x_j$ in $L_i(A)$ is given as linear combination of Lie monomials.

Although $P_1, \ldots, P_k$ are unique, their decomposition into sum of monomials is not. The reader must notice that our algorithm will start with the given of the d_i's, the x_j's and λ_j's as inputs and that it avoids any calculation in the free Lie algebra. Here we are not concerned with the problem of the decomposition of Lie polynomials in some particular basis. For instance there is no *a priori* bound for the d_i's (an upper bound could have been given by the dimension of $L_i(A)$ that satisfies the Witt formula: $\sum_{d|n} d \dim(L_d(A)) = |A|^n$).

Under these hypothesis, our algorithm gives a kth-order approximation $\varphi_k(P)$. Let $l_k(P)$ be the number of elementary factors in $\varphi_k(P)$. We will give an upper bound of $l_k(P)$ with respect to $d_1, \ldots, d_k$. We give this estimation by following the step numbers of Part 2.5.

- For $a \in A$ and $\lambda \in \mathbb{R}$, we have obviously $l_k(\lambda a) = 1$ (see step 1).
- If x_n is a monomial of length n, then, with the notations of step 2, we have:

  $$l_n(\lambda x_n) \leq 2(l_p(\lambda a_p) + l_q(b_q)).$$

  By induction on n we see that

  $$l_n(\lambda x_n) \leq 3 \times 2^{n-1} - 2.$$

- For an homogeneous Lie polynomial $P_i = \sum_{j=1}^{d_i} \lambda_j x_j \in L_i(A)$, step 3 implies that

  $$l_i(P_i) \leq 3d_i \times 2^{i-1}.$$

- Therefore, we have, by using step 4:

  $$l_k(P_i) \leq 3^{k-i+1} d_i \ 2^{i-1} = 3^k \left(\frac{2}{3}\right)^{i-1} d_i. \tag{7}$$

– Finally, we get

$$l_k(P) \leq 3^k \sum_{i=1}^{k} \left(\frac{2}{3}\right)^{i-1} d_i \leq 3^{k+1} \max\{d_i\}. \tag{8}$$

Remark. — If $P = a + b$ it is known (see [14, 4]) that $l_{2k}(a+b) = 2^{k+1} - 1$ for $k \leq 4$, that is an exponential bound.

We know better estimates in some particular cases. If $P = 0$, it has been shown ([2]) that $k \leq l_k(P) \leq k^2$, that is a polynomial bound. It implies that for any k, there exists a Lie polynomial $P_k \in L_k(A)$ such that $\varphi_k(P_k)$ is a product of less than k^2 factors. This is also far less than the sum of dimensions $\sum_{i=1}^{k} \dim(L_i(A))$.

These two examples and example 2.6 show that the minimal number of factors of $\varphi_k(P)$ depends on the polynomial P and on its decomposition into monomials. In this paper we will not seek minimal approximants (this question is considered in [5, 14] and in the remark at the end of Part 3.3).

3.2 Improvements

Our construction is certainly not optimal. We can then improve it a lot. Let us notice that the steps of the construction are independent of each other. Each step can then be improved separately. We will discuss now about some possible modifications (we refer to the step numbers of Part 2.5).

We present first a simple improvement of step 4. When k and n have not the same parity, the recursion formula for $\varphi_k(P)$, can be replaced by:

$$\begin{aligned} &\text{if } n \text{ is even and } k \text{ odd, then } \varphi_k(P) = \varphi_{k-1}(P)[2^{-1/n}]\ \varphi_{k-1}(P)[-2^{-1/n}], \\ &\text{if } n \text{ is odd and } k \text{ even, then } \varphi_k(P) = \varphi_{k-1}(P)[2^{-1/n}]\ \varphi_{k-1}(P)[-2^{-1/n}]^{-1}. \end{aligned} \tag{9}$$

This formula allows to reduce the theoretical bound (7) (and then the bound (8)) for the size of the approximation, since it has two factors instead of three. When n and k have the same parity, there is no such formula with two factors.

We are now interested in step 2. We have chosen in our algorithm to write λx as $[\lambda a, b]$. The idea was: if λ is an integer, then the elementary factors of $\varphi_n(\lambda x)$ will use only integer. However we can see on examples that, with this construction, there are no cancellations between the factors of $\varphi_n(\lambda x)$ (we call cancellation the occurrence of a product $\exp(y)\exp(-y)$). For example, we get

$$\varphi_2([x,y]) = \exp(x)\exp(y)\exp(-x)\exp(-y)$$

and we will test our algorithm on $\varphi_4(\lambda[[x,y],[x,y]])$. For $\lambda = 1$, algorithm produces

$$\varphi_4([[x,y],[x,y]]) = 1$$

and for $\lambda = 2$, it gives

$$\varphi_4(2[[x,y],[x,y]]) = \exp(2x)\exp(y)\exp(-x)\exp(-y)\exp(-x)\exp(y)$$
$$\exp(2x)\exp(-y)\exp(-x)\exp(y)\exp(-x)\exp(-y).$$

Thus we see that the cancellations are not preserved by the algorithm.

To avoid this kind of problem we can use another method. We first construct $\varphi_n(x)$. Then, if x is a Lie monomial of length n, we get

$$\varphi_n(|\lambda|x) = \varphi_n(x)[|\lambda|^{1/n}].$$

Finally we have $\varphi_n(\lambda x) = \varphi_n(|\lambda|x)^\sigma$, where $\sigma = \pm 1$ is such that $\lambda = \sigma|\lambda|$. This formula allows to keep cancellations from $\varphi_n(x)$ to $\varphi_n(\lambda x)$.

This construction does not give theoretical improvement for the size of the approximation: it does not reduce the bounds (7) and (8). But in practice, it gives often an approximation with less factors. On the other hand, it makes use of algebraic numbers even if λ is an integer.

3.3 Example

Let $P = a + [b,a] + [a,[b,a]]$. Let us compute the approximation $\varphi_3(P)$.

We first show how to get the expansion (6) of $\exp(P)$ at the order 3.

$$\begin{aligned}\exp(P) &= \exp(a + [b,a] + [a,[b,a]])\\ \exp(-a)\exp(P) &= \exp([b,a] + [a,[b,a]] - \tfrac{1}{2}[a,[b,a]] + R_{\geq 4})\\ \exp(-[b,a])\exp(-a)\exp(P) &= \exp(\tfrac{1}{2}[a,[b,a]] + R'_{\geq 4}) \qquad (10)\end{aligned}$$

We thus deduce that $P_1 = a$, $P_2 = [b,a]$, $P_3 = \frac{1}{2}[a,[b,a]]$.

Algorithm starts now by computing $\varphi_3(P_i)$.

- We get from the previous example (Part 2.6)

  $$\varphi_3(P_1) = \exp(a),\ \varphi_2(P_2) = \exp(b)\exp(a)\exp(-b)\exp(-a).$$

  From Formula (9) we obtain $\varphi_3(P_2) = \varphi_2(P_2)[\frac{1}{\sqrt{2}}]\varphi_2(P_2)[\frac{-1}{\sqrt{2}}]$ and then:

  $$\varphi_3(P_2) = \exp(\tfrac{1}{\sqrt{2}}b)\exp(\tfrac{1}{\sqrt{2}}a)\exp(\tfrac{-1}{\sqrt{2}}b)\exp(\tfrac{-1}{\sqrt{2}}a)\exp(\tfrac{-1}{\sqrt{2}}b)\exp(\tfrac{-1}{\sqrt{2}}a)$$
  $$\exp(\tfrac{1}{\sqrt{2}}b)\exp(\tfrac{1}{\sqrt{2}}a)$$

- For P_3, we have $\varphi_3(P_3) = \varphi_1(\frac{1}{2}a)\varphi_2([b,a])\varphi_1(\frac{1}{2}a)^{-1}\varphi_2([b,a])^{-1}$, that is:

  $$\varphi_3(P_3) = \exp(\tfrac{1}{2}a)\exp(b)\exp(a)\exp(-b)\exp(-a)\exp(-\tfrac{1}{2}a)\exp(a)\exp(b)$$
  $$\exp(-a)\exp(-b).$$

  Noticing that $\exp(-a)\exp(-\frac{1}{2}a)\exp(a) = \exp(-\frac{1}{2}a)$ we get:

  $$\varphi_3(P_3) = \exp(\tfrac{1}{2}a)\exp(b)\exp(a)\exp(-b)\exp(-\tfrac{1}{2}a)\exp(b)\exp(-a)\exp(-b).$$

The approximation $\varphi_3(P)$ is then given by $\varphi_3(P) = \varphi_3(P_1)\varphi_3(P_2)\varphi_3(P_3)$. Taking together the end term of $\varphi_3(P_2)$ and the first term of $\varphi_3(P_3)$, we get:

$$\varphi_3(P) = \exp(a)\exp(\tfrac{1}{\sqrt{2}}b)\exp(\tfrac{1}{\sqrt{2}}a)\exp(\tfrac{-1}{\sqrt{2}}b)\exp(\tfrac{-1}{\sqrt{2}}a)\exp(\tfrac{-1}{\sqrt{2}}b)\exp(\tfrac{-1}{\sqrt{2}}a)$$
$$\exp(\tfrac{1}{\sqrt{2}}b)\exp(\tfrac{1+\sqrt{2}}{\sqrt{2}}a)\exp(b)\exp(a)\exp(-b)\exp(-\tfrac{1}{2}a)\exp(b)\exp(-a)\exp(-b).$$

Remark. — As pointed out at the end of Part 3.1, we could find a shorter solution. If we calculate $\exp(Ab)\exp(Ba)\exp(Cb)\exp(Da)\exp(Eb) = \exp(P)$, we find up to order 3:

$$\begin{aligned}
&P = (D+B)\ \ a + (E+C+A)\ \ b + \\
&+\tfrac{1}{2}\left((A-C)B - (D+B)E + (A+C)D\right)[a,b] \\
&+\tfrac{1}{12}\left((D+B)^2E + (A+C)D^2 + (2AB-4BC)D + B^2C + AB^2\right)[a,[a,b]] \\
&+\tfrac{1}{12}\left((B+D)E^2 + (-4(A+C)D + 2BC - 4AB)E + (A+C)^2D\right)[[a,b],b].
\end{aligned}$$

We thus find a five-factors third-order approximant for $P = a + [b,a] + [a,[b,a]]$, by setting

$$B = \tfrac{1}{2}(1+\varepsilon\sqrt{65}),\ D = \tfrac{1}{2}(1-\varepsilon\sqrt{65}), C = -\tfrac{1}{8},\ E = \tfrac{B+4}{8},\ A = -\tfrac{B+3}{8}.$$

This method makes use of calculations in the free algebra and requires to solve polynomial system. It could not be so easely generalized to higher orders.

4 Application to Control Theory

The use of our approximations in control theory (in particular for the motion planning problem) will be discussed in a next paper. We give here an example of such application.

Let $X_1, \ldots, X_m$ be vector fields on $\mathbb{R}^n$ and (Σ) the control system

$$\dot{x} = \sum_{i=1}^{m} u_i(t) X_i(x).$$

Let $\mathcal{L}(X_1, \ldots, X_m)$ be the Lie algebra generated by $X_1, \ldots, X_m$. For a given X in $\mathcal{L}(X_1, \ldots, X_m)$, we consider the problem of "approximating" a point lying on the flow of X. More precisely, the problem is to find a trajectory steering the system from a point p to a point $\exp(tX + o(t^k))p$ ($o(t^k)$ denotes here a vector field with a norm in $o(t^k)$).

Any composition of flows of the X_i's is a trajectory of the system (Σ). The end-point of such a trajectory is written $\exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s})p$. Therefore sequences λ_j and X_{i_j} such that

$$\exp(\lambda_1 X_{i_1}) \cdots \exp(\lambda_s X_{i_s})p = \exp(tX + o(t^k))p \tag{11}$$

give a solution to our problem.

We first translate the problem in the context of the free Lie algebras. Let $L(A)$ be the free Lie algebra generated by an alphabet A of m elements. We denote by $o(t^k)$ an element $t^k R(t)$ in $\hat{L}(A)$, with $R(t) \in \hat{L}(A)$ for each t, and $R(0) = 0$. Using the universal Lie algebra morphism $\phi : L(A) \to \mathcal{L}(X_1, \ldots, X_m)$, $a_i \mapsto X_i$, any relation in the free Lie algebra like

$$\exp(\lambda_1 a_1) \cdots \exp(\lambda_s a_s) = \exp(tP + o(t^k)) \tag{12}$$

will give a relation (11) in $\mathcal{L}(X_1, \ldots, X_m)$.

That is why we will establish universal approximations (12) in $L(A)$. Next example shows how it can be done.

4.1 Example

Let $P = a + [b, a] + [a, [b, a]]$ (the same polynomial as in Part 3.3) and let us try to build an approximation in $o(t)$. We first compute the expansion (6) of $\exp(tP)$ at order 3. Formula (10) gives

$$\exp(-t[b, a]) \exp(-ta) \exp(tP) = \exp((t - \tfrac{t^2}{2})[a, [b, a]] + R'_{\geq 4}(t))$$

where $R'_{\geq 4}(t) = t^2 R'(t)$. We thus deduce that

$$\exp(tP) = \exp(t\, a) \exp(t\, [b, a]) \exp(t\, [a, [b, a]]) \exp(o(t)).$$

Thus the expansion (6) at order 3 gives only an approximation "in $o(t)$". In order to get an approximation "in $o(t^2)$", we should go until sixth-order in decomposition 6.

We would like to convince the reader that

$$\alpha = \varphi_1(a)[t]\, \varphi_2([b, a])[t^{1/2}]\, \varphi_3([a, [b, a]])[t^{1/3}]$$

is an approximation of $\exp(tP)$ in the form (12).

This construction could be generalized to any Lie polynomial P. We would have to refine Lemma 9 and then use the approximations $\varphi_k(P)$ for the homogeneous Lie polynomials. This will be detailed in a next paper.

5 References

[1] **Bourbaki, N.**, *Groupes et algèbres de Lie*, Éléments de Mathématiques, Hermann, Paris, 1972

[2] **Falbel, E., Koseleff, P.-V.**, *Parallelograms*, Preprint (1996)

[3] **Jacob, G.**, *Motion Planning by piecewise constant or polynomial inputs*, Proceedings of the IFAC Nonlinear Control Systems Design Symposium (1992)

[4] **Koseleff, P.-V.**, *Relations among Formal Lie Series and Construction of Symplectic Integrators*, AAECC'10 proceedings, Lect. Not. Comp. Sci. **673** (1993)

[5] **Koseleff, P.-V.**, *Exhaustive Search of Symplectic Integrators Using Computer Algebra*, Fields Institute Communications **10** (1996)

[6] **Lafferriere, G., Sussmann H.**, *Motion Planning for controllable systems without drift*, Proceedings of the 1991 IEEE International Conference on Robotics and Automation (1991)

[7] **Laumond, J.P.**, *Nonholonomic Motion Planning via Optimal Control*, Algorithmic Foundations of Robotics (1995)

[8] **MacLachlan, R. I.**, *On the numerical integration of ordinary differential equations by symmetric composition methods*, SIAM J. Sci. Comp. **16(1)** (1995), 151–168

[9] **Magnus** *et al.*, *Combinatorial Group Theory: Presentation of Groups in Terms of Generators and Relations*, J. Wiley & Sons, 1966

[10] **Reutenauer, C.**, *Free Lie algebras*, Oxford Science Publications, 1993

[11] **Steinberg, S.**, *Lie Series, Lie Transformations, and their Applications*, in *Lie Methods in Optics*, Lec. Notes in Physics **250** (1985)

[12] **Suzuki, M.**, *General Theory of higher-order decomposition of exponential operators and symplectic integrators*, Physics Letters A **165** (1992), 387–395

[13] **Suzuki, M.**, *General nonsymetric higher-order decompositions of exponential operators and symplectic integrators*, Physic Letters A **165** (1993), 387–395

[14] **Yoshida, H.**, *Construction Of Higher Order Symplectic Integrators*, Ph. Letters A **150,** (1990), 262–268

This article was processed using the LaTeX macro package with LLNCS style

The Number of Sides of a Parallelogram

FALBEL, E., KOSELEFF, P.-V.
Discrete Mathematics and Theoretical Computer Science (DMTCS) 3 :2 (1999), pp 33-42

Discrete Mathematics and Theoretical Computer Science **3**, 1999, 33–42

The Number of Sides of a Parallelogram

Elisha Falbel and Pierre-Vincent Koseleff

Institut de Mathématiques, Université Paris 6, Case 82
4 place Jussieu, F-75252 Paris Cedex 05.
email: {falbel,koseleff}@math.jussieu.fr

received 13 Oct 97, revised 3 Nov 1998, accepted 10 Jan 1999.

We define parallelograms of base a and b in a group. They appear as minimal relators in a presentation of a subgroup with generators a and b. In a Lie group they are realized as closed polygonal lines, with sides being orbits of left-invariant vector fields. We estimate the number of sides of parallelograms in a free nilpotent group and point out a relation to the rank of rational series.

Keywords: Lie algebras, free group, Magnus group, lower central series, Lyndon basis

1 Introduction

In $\mathbb{R}^2$ a parallelogram of base a and b can be defined as a closed polygon with the minimum number of sides parallel to a and b. In that paper we also consider parallelograms defined in more general groups.

In section **1.** we first give some definitions and examples of parallelograms in Lie groups. These examples show the various complex situations occurring in the general case. In this paper we concentrate our attention on free nilpotent groups. This analysis will give universal properties for parallelograms. We obtain

Theorem. The number of sides of a parallelogram on a free nilpotent group on two generators of order n is between n and n^2.

We do not know what is the exact number of sides of parallelograms in a free nilpotent group neither how many *non-equivalent* parallelograms exist. We hope that an investigation of parallelograms might help understand general nilpotent groups. In particular it will be interesting to find presentations with relators of minimal size.

We have chosen in this paper to recall the basic properties and constructions of free Lie algebras in order to make it self-contained. That is done in section **2.** In the last section we then introduce mth-order parallelograms and prove our result. A connection with rational series is pointed out at the end of the paper.

Our initial motivation to study parallelograms was the notion of curvature and holonomy of a connection for Riemannian manifolds and the generalization of those notions to sub-Riemannian geometry (see [FGR] and [BeR]). In classical differential geometry, curvature appears as the quadratic term in the asymptotic expansion of holonomy around short (four-sided) parallelograms, holonomy being the measure of the difference of the vector field by parallel translation around a closed loop. In the case of sub-Riemannian manifolds, the tangent space is naturally a nilpotent group ([BeR]) and the holonomy associated to it will be calculated using *parallelograms* with many sides. The analog of sectional curvatures should be the holonomy associated to different parallelograms.

Another motivation is the approximation of a given element of the group by elements of a given subgroup. This occur for example in the search of symplectic integrators (see [K, Su]) that give numerical schemes for long-time integration of hamiltonian systems. Namely we try to approximate $\exp(x+y)$ by a product of $\exp(x)$ and $\exp(y)$. In this frame, minimal length of mth-order approximants are bounded by approximately 2^m.

Acknowledgements

Authors would like to thank G. Duchamp for fruitful discussions and many helpful comments. He did focus our attention to the theory of noncommutative series. The first author would like to thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for financial support and the *Institut de Mathématiques (UMR 7586)* for its kind support.

2 Definitions and examples

Definition 2.1 *A* segment *in a Lie group is a curve obtained by following the orbit of left-invariant vector field. It has initial and end points. Two segments are parallel if they are orbits of two dependent left-invariant vector fields.*

Definition 2.2 *A* polygonal line *in a Lie group is a curve obtained by concatenation of segments, two consecutive segments being not parallel. This is a sequence of segments where the end point of one of them coincides with the initial point of its successor. Each segment is called a* side.

Observe that once we have fixed a left invariant vector field X, a side is of the form $\gamma(t) = x_0 \exp(t\lambda X)$, where $0 \le t \le 1$. In that case we call $|\lambda|$ the *length* of the side. $\gamma(0)$ is its initial point and $\gamma(1)$ its end point.

Definition 2.3 *A* polygon *in a Lie group is a closed polygonal line. Its* length *is the sum of its sides lengths.*

Definition 2.4 *A* parallelogram *of* base X *and* Y *in a Lie group is a polygon with sides of integer length, obtained from the two given left-invariant vector fields* X *and* Y, *with minimum length. Two parallelograms are* equivalent *if there exists a group isomorphism which maps one parallelogram onto the other.*

In order to describe explicitly a polygonal line with n sides, let $\mathcal{F} = \{X_\alpha\}$ be a family of linearly independent vectors in the Lie algebra $\mathbf{g}$ of the Lie group G. Fix $x_0 = 1 \in G$. We write $\gamma_j(t) = x_{j-1}\exp(t\lambda_j X_{\alpha_j})$ for $x_j = x_{j-1}\exp(\lambda_j X_{\alpha_j})$, $0 \le t \le 1$ and $1 \le j \le n$. Here we require that X_{α_j} and $X_{\alpha_{j+1}}$ are independent. Denote by $P(\lambda_1 X_{\alpha_1}, \ldots, \lambda X_{\alpha_n})$ the polygonal line defined in this way.

Example 2.1 *Consider the abelian Lie group* $\mathbb{R}^n$. *A parallelogram in that group is clearly a* parallelogram.

Example 2.2 *Consider the Heisenberg group H^3 with Lie algebra generated by X,Y,Z, with $[X,Y]=Z$, all other brackets being null. One can verify, using the Campbell-Hausdorff formula that both*

$$P_8(X,Y) = P(X,Y,-X,-2Y,-X,Y,X) \text{ and } P_8'(X,Y) = P(X,Y,-X,-Y,-X,Y,X,Y)$$

are parallelograms. They are not equivalent as P_8 has at least one side of length two. On the other hand starting with X,Z we get a parallelogram of 4 sides.

Example 2.3 *Let L^4 be a free nilpotent group of order 4, generated by X and Y. We can verify that $P(X,Y,-2X,-Y,X,Y,X,-Y,-2X,Y,X,-Y)$ is a parallelogram. It has length 14. An interesting question would be to know all non-equivalent parallelograms.*

Example 2.4 *If the group generated by $\exp(X)$ and $\exp(Y)$ is free, then there is no parallelogram of base X and Y.*

We thank the referee for pointing out the two following examples.

Example 2.5 *As a result of a theorem by* SANOV *([Sa]), for $X^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $X^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, the group $G =< \exp(2X^+), \exp(2X^-) >$ is free (see also [LS]), so there exist no parallelogram of base $2X^+$ and $2X^-$. Moreover it is straightforward that $P = (\exp(X^+)\exp(-X^-))^6 = 1$ is a parallelogram of length 12 with base X^+ and X^-.*

We could have given a more general definition of a parallelogram in an arbitrary group. Let a and b be two elements on a group G and $G{<}a,b{>}$ be the subgroup generated by a,b. Consider the set of all relators, i. e., the set of words in a,b,a^{-1},b^{-1} which are the identity in G. One should consider only reduced words in the sense that if a is of order n and a^n appears in a word, one should substitute the identity for a^n. The same for b. A *parallelogram* of base a,b is a reduced relator (in the above sense) of minimal length with letters a,b,a^{-1},b^{-1}. Of course if $G{<}a,b{>}$ is free in a,b there is no parallelogram.

Example 2.6 *In the case of the symmetric group*

$$\mathbf{S}_3 =< \sigma_1, \sigma_2;\ \sigma_i^2 = 1, \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 >$$

one can verify that a minimal relator with base σ_1,σ_2 is $(\sigma_1\sigma_2)^3$ of length 6. On the other hand we have also

$$\mathbf{S}_3 =< \sigma_1, \sigma_3 = \sigma_2\sigma_1;\ \sigma_1^2 = 1, \sigma_3^3 = 1, \sigma_1\sigma_3 = \sigma_3\sigma_1 >$$

that has a minimal relator of length 4.

In the case of Lie groups we would like to define *infinitesimal parallelograms*, that is parallelograms which remain the same in form when their sides are changed by a conformal factor. They will not exist in general but in the case of graded nilpotent groups their existence is assured.

Example 2.7 *Consider the Lie group with Lie algebra generated by X,Y with $[X,Y]=X$. Then we can construct a parallelogram which is not infinitesimal. Observe that*

$$\exp(tY)\exp(uX)\exp(-tY) = \exp(u\exp(-t)X)).$$

So if $t = -\ln 2$ *and* $u = 1$, *we have* $\exp(tY)\exp(uX)\exp(-tY)\exp(-u\exp(-t)X) = 1$. *That is a parallelogram of length 5 with base* $\ln 2\, Y$ *and* X. *It is clear that if we change the sides by a conformal factor this will no longer be a parallelogram. More generally, a polygon is a product*

$$\exp(c_1Y)\exp(d_1X)\cdots\exp(c_nX)\exp(d_nY)$$

such that $\sum_i(\sum_{j\leq i} c_j)\exp(-d_i) = 0$. *The previous equation has clearly no integer solutions.*

Example 2.8 *Let us consider in* $\mathbb{R}^2$, $X = \partial_x$ *and* $Y = f(x)\partial_y$ *for a given analytic function* f. *The Lie algebra* $L(X,Y)$ *is in general infinite dimensional as* $(\mathrm{ad}\ X)^nY = f^{(n)}(x)\partial_y$ *and is spanned, as vector space by* X *and* $\{(\mathrm{ad}\ X)^nY\}$. *By noticing that* $\exp(\lambda\mathrm{ad}\ X)Y = f(x+\lambda)\partial_y$, *we deduce that*

$$\exp(tX)\exp(uY)\exp(-tX) = \exp(uf(x+t)\partial_y)$$

so

$$\begin{aligned}\exp(X)\exp(Y)\exp(-X)\exp(Y)\exp(X)\exp(-Y)\exp(-X)\exp(-Y) &= \\ \exp(f(x+1)\partial_y)\exp(f(x)\partial_y)\exp(-f(x+1)\partial_y)\exp(-f(x)\partial_y) &= 1.\end{aligned}$$

This gives a parallelogram of length 8.

3 Magnus Groups and Algebras

Let us first introduce some notations and recall some results about free groups, free associative algebras and free Lie algebras. All these results can be found in ([B, La, R]).

Let X be a set (alphabet). We denote by X^* the free monoid generated by X, that is, the set of words including the empty word denoted by 1, with concatenation as a product. X^* is totally ordered by the lexicographic order. The free magma $M(X)$ is the set of words with parentheses, generated by X and $A(X)$ denotes the free associative algebra, that is to say the $\mathbb{Q}$-algebra of X^*. An element P in $A(X)$ will be written $\sum_{w\in X^*}(P,w)w$.

We denote by $L(X)$ the free Lie algebra on A. It is the quotient of the $\mathbb{Q}$-algebra of $M(X)$ by the ideal generated by the elements (u,u) and $(u,(v,w)) + (v,(w,u)) + (w,(u,v))$. The associative algebra $A(X)$ may be identified to the enveloping algebra of $L(X)$ by considering $[v,w] = vw - wv$. We denote by ad x the map $y \mapsto [x,y]$.

The free group generated by X is denoted by $F(X)$.

3.1 Gradations

The sets $L(X), F(X)$ so as $A(X)$ are graded by

— the length (the unique homomorphism that extends the function $x \mapsto 1$ on X). For $x \in X^*$ (resp. $F(X), M(X)$) $|x|$ denotes the length. $L_n(X)$ (resp. $A_n(X)$) is the submodule generated by monomials of length n.

— the multi-degree which is the unique homomorphism from X^* (resp. $F(X), M(X)$) onto $\mathbb{N}^{(X)}$ that extends $x \mapsto \mathbb{1}_x$. For a given α in $\mathbb{N}^{(X)}$, $L^\alpha(X)$ (resp. $A^\alpha(X)$) denotes the submodule generated by monomials of degree α.

Definition 3.1 *Let A,B be subgroups of a group C. We denote by (A,B) the set of all commutators $(a,b) = aba^{-1}b^{-1}$. Starting with $F_{\geq 1}(X) = F(X)$ and defining $F_{\geq n}(X) = (F_{\geq 1}(X), F_{\geq n-1}(X))$, we get the so-called* lower central series.

As a consequence, we have $(F_{\geq n}(X), F_{\geq m}(X)) \subset F_{\geq n+m}(X)$ and $F(X)/F_{\geq n}(X)$ is an abelian group.

3.2 Formal series

We define $\hat{L}(X)$ and $\hat{A}(X)$ as $\hat{L}(X) = \prod_{n\geq 0} L_n(X)$ $\hat{A}(X) = \prod_{n\geq 0} A_n(X)$. We will write $x \in \hat{L}(X)$ (resp. $\hat{A}(X)$) as a series $\sum_{n\geq 0} x_n$. $\hat{L}(X)$ so as $\hat{A}(X)$ are algebras with multiplications law

$$(xy)_n = \sum_{p+q=n} x_p y_q, \ ([x,y])_n = \sum_{p+q=n} [x_p, y_q]. \tag{1}$$

We will also use $\hat{L}_{\geq p}(X) = \prod_{n\geq p} L_n(X)$ $\hat{A}_{\geq p}(X) = \prod_{n\geq p} A_n(X)$. The set $\Gamma(X) = 1 + \hat{A}_{\geq 1}(X)$ is called the *Magnus group*. It is a subgroup of the invertible elements of $\hat{A}(X)$. One defines the exponential and the logarithm as

$$\begin{array}{rcl} \exp : \hat{A}_{\geq 1}(X) & \to & \Gamma(X) \\ x & \mapsto & \sum_{n\geq 0} \dfrac{x^n}{n!}, \end{array} \qquad \begin{array}{rcl} \log : \Gamma(X) & \to & \hat{A}_{\geq 1}(X) \\ x & \mapsto & -\sum_{n\geq 1} \dfrac{(1-x)^n}{n}. \end{array}$$

They are mutually reciprocal functions and we have (see [B, Ch. II, §5]) the

Theorem 3.1 (Campbell-Hausdorff) *For $x,y \in \hat{L}_{\geq 1}(X)$,*

$$H(x,y) = \log[\exp(x)\exp(y)] \in \hat{L}_{\geq 1}(X). \tag{2}$$

Denoting by $\hat{E}_{\geq n}(X) = \exp(\hat{L}_{\geq n}(X))$, we get

Corollary 3.1 *The set $\hat{E}_{\geq 1}(X) = \exp(\hat{L}_{\geq 1}(X)) \subset \Gamma(X)$ is a group.*

$\hat{E}_{\geq 1}(X)$ acts on itself by conjugacy and we have $\exp(x)\exp(y)\exp(-x) = \exp(\exp(\mathrm{ad}\, x)y)$.

Definition 3.2 *Let us consider the Magnus map $\mu : F(X) \to \Gamma(X)$ as the unique group homomorphism that extends $x \mapsto 1 + x$, for $x \in X$. We set $D_{\geq n}(X) = \mu^{-1}(1 + \hat{A}_{\geq n}(X))$. This is Magnus' n-th dimension subgroup of F.*

Definition 3.3 *Let us consider the map $\mu' : F(X) \to \Gamma(X)$ as the unique group homomorphism that extends $x \mapsto \exp(x)$, for $x \in X$. We set $D'_{\geq n}(X) = \mu'^{-1}(1 + \hat{A}_{\geq n}(X))$.*

This defines central filtrations of $F(X)$. We have clearly that $F_{\geq n}(X) \subset D_{\geq n}(X)$ and $F_{\geq n}(X) \subset D'_{\geq n}(X)$. In fact Magnus proved a stronger result (see [B])

Proposition 3.1 $D_{\geq n}(X) = D'_{\geq n}(X) = F_{\geq n}(X)$

Let $N_n(X)$ be the free nilpotent group of class n (or order $n+1$) on X. That is

$$1 \to F_{\geq n+1}(X) \to F(X) \to N_n(X) \to 1 \tag{3}$$

We will use the following corollary to establish the lower bound to the number of sides of parallelogram on the free nilpotent group.

Corollary 3.2 *The projection of g in $F(X)$ onto $N_n(X)$ is the identity if and only if $\mu'(g) \in \hat{E}_{\geq n}(X)$.*

In fact we need only the if part of the corollary for the lower bound, that is not dependent on Magnus result but on the inclusion $F_{\geq n}(X) \subset D_{\geq n}(X)$.

4 *m*th-order parallelograms

Definition 4.1 *The* order *of g in $F(X)$ is the biggest integer k such that $g \in F_{\geq k}(X)$. An element of order k will be called* kth-order polygon.

Using proposition (3.1), a mth-order polygon g satisfies

$$\begin{aligned} g &= x^{a_1}y^{b_1}\cdots x^{a_n}y^{b_n} \in F_{\geq m}(X), &(4)\\ \mu'(g) &= \exp(a_1x)\exp(b_1y)\cdots\exp(a_nx)\exp(b_ny) \in 1+\hat{A}_{\geq m}(X), &(5)\\ \mu(g) &= (1+x)^{a_1}(1+y)^{b_1}\cdots(1+x)^{a_n}(1+y)^{b_n} \in 1+\hat{A}_{\geq m}(X). &(6) \end{aligned}$$

Here none of $a_i's$ nor $b_i's$ is 0.

Definition 4.2 *The* length $l : F(X) \to \mathbb{N}$ *is the unique homomorphism that extends $x \mapsto 1, x^{-1} \mapsto 1$, for x in X. If $g = x_1^{i_1}\cdots x_p^{i_p} \in F(X)$, we sill say that it is a p*-sided *polygon. For example $xyx^{-1}y^{-1}$ is a 4-sided second-order parallelogram of length 4. In formula (4), we have $l(g) = \sum_{i=1}^{n}(|a_i|+|b_i|)$.*

We thus deduce that for any g_1, g_2 in $F(X)$, we have $l(g_1g_2) \leq l(g_1)+l(g_2)$. The inequality is strict only if terms of g_1 cancel terms of g_2.

Definition 4.3 *For $m \in \mathbb{N}$, we define l_m as the lowest length of mth-order polygons. A* mth-order parallelogram *will be a mth-order polygon of minimal length.*

Before discussing the lower and upper bounds for the length and the number of factors of mth-order parallelograms, let us show some transformations that preserve polygons.

Proposition 4.1 *Let $\alpha\beta$ be a mth-order polygon then so is $\beta\alpha$.*

Corollary 4.1 *If g is a $(2p+1)$-sided mth-order polygon then there exists a $2p$-sided mth-order polygon.*

Proof. — The proposition comes from the fact that $F/F_{\geq m}(X)$ is abelian.
Let us suppose that $g = x^{a_1}y^{b_1}\cdots y^{b_p}x^{a_{p+1}}$ is a mth-order polygon. Then

$$x^{(a_1+a_{p+1})}y^{b_1}\cdots y^{b_p} \tag{7}$$

has smaller length as $|a_1+a_{p+1}| \leq |a_1|+|a_{p+1}|$ and is also a mth-order polygon. □

We can now suppose that for any integer m, an mth-order parallelogram has an even number of factors. We will now discuss lower and upper bound of l_m.

4.1 Lower bound

Proposition 4.2 *For any $m \in \mathbb{N}$ we have $m \leq l_m$.*

Proof. — Let us consider the following equality

$$\exp(a_1x)\exp(b_1y)\cdots\exp(a_nx)\exp(b_ny) = \exp(z). \tag{8}$$

where $z \in \hat{L}_{\geq m}(X)$ and none of the $a_i's$ nor b_i's is 0. Considering the word $w = (xy)^n$, we have

$$(\exp(z), w) = \prod_{i=1}^{n} a_ib_i \neq 0$$

and so $m \leq 2n \leq l_m$. In fact the number of sides itself is bigger than m. □

4.2 Upper bound

First of all, let us show some small-order parallelograms.

If $m = 1$ $g_1 = x$ or $g_1 = y$ is convenient. If $m = 2$, we find $g_2 = xyx^{-1}y^{-1}$ thus $l_2 \le 4$. In fact $l_2 = 4$ which is a consequence of the following

Lemma 4.1 *For any $m \ge 2$, l_m is even.*

Proof. — This is a consequence of

$$\mu(g) = (1+x)^{a_1}(1+y)^{b_1}\cdots(1+x)^{a_1}(1+y)^{b_1} = 1 + (a_1 + \cdots + a_n)x + (b_1 + \cdots + b_n)y \ (\hat{A}_{\ge 2}(X)).$$

So if $\mu(g)$ belongs to $\hat{A}_{\ge m}(X)$ we have $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i = 0$ thus $\sum |a_i|$ and $\sum |b_i|$ are even. □

We have seen g_2 as the commutator of two first-order polygons. We will now build a sequence g_m of mth-order polygons, each g_m being constructed as commutator of g_p and g_{m-p} for some p. We first use the following lemma

Lemma 4.2 *Let g_p and g_q be two polygons of order p and q respectively, then (g_p, g_q) has order at least $p+q$ and has length at most $2(l(g_p)+l(g_q))$.*

Remark. — This is also a consequence of the fact that $(F_{\ge n}(X))_n$ is a central filtration but we will show it by using the Hausdorff series.

Proof. — Let us write

$$P_p = \mu'(g_p) = \exp(x) = \exp(\textstyle\sum_{k\ge p} x_k), P_q = \mu'(g_q) = \exp(y) = \exp(\sum_{k\ge q} y_k). \tag{9}$$

then we have

$$P_p P_q P_p^{-1} P_q^{-1} = \exp(\exp(\mathrm{ad}\, x)y)\exp(-y) = \exp(H(\exp(\mathrm{ad}\, x)y, -y)). \tag{10}$$

But

$$\begin{aligned} H(\exp(\mathrm{ad}\, x)y, -y) &= H_1(\exp(\mathrm{ad}\, x)y, -y) + \sum_{k\ge 2} H_k(\exp(\mathrm{ad}\, x)y, -y) && (11)\\ &= \exp(\mathrm{ad}\, x))y - y + \textstyle\sum_{k\ge 2} H_k(\exp(\mathrm{ad}\, x)y, -y) && (12)\\ &= [x,y] + \sum_{k\ge 2}\frac{1}{k!}(\mathrm{ad}\, x)^k y + \sum_{k\ge 2} H_k(\exp(\mathrm{ad}\, x)y, -y). && (13) \end{aligned}$$

But $(\mathrm{ad}\, x)^k y \in \hat{L}_{\ge kp+q}(X) \in \hat{L}_{\ge 2p+q}(X)$ and $H_k(\exp(\mathrm{ad}\, x)y, -y) = H_k(\exp(\mathrm{ad}\, x)y - y, -y) \in \hat{L}_{\ge p+2q}(X)$. In conclusion, if $[x_p, y_q] \ne 0$, then $g_{p+q} = (g_p, g_q)$ is a $p+q$-th order polygon and has length $2(l(g_p)+l(g_q))$. In order to be sure to obtain a $(p+q)$-th order polygon let us show that

Lemma 4.3 *Let $\alpha \in F_{\ge p}(X)$ and $\beta \in F_{\ge q}(X)$ such that*

$$\mu'(\alpha) = \exp(x) = \exp(\textstyle\sum_{k\ge p} x_k), \mu'(\beta) = \exp(y) = \exp(\sum_{k\ge q} y_k). \tag{14}$$

If x_p and y_q are not proportional, then (α, β) has order exactly $p+q$.

Proof. — This is a consequence of the following lemma.

Lemma 4.4 *Let $x_p \in L_p(X)$ and $y \in L_q(X)$. If x_p and y_q are not proportional then $[x_p, y_q] \neq 0$.*

Proof. — Let us write

$$x_p = \sum_{w \in X_p^*} (x_p, w) = \sum_{i=1}^{n} \lambda_i w_i, y_q = \sum_{w \in X_q^*} (y_q, w) = \sum_{i=1}^{n'} \lambda'_i w'_i. \tag{15}$$

Here we have $w_i < w_j$ if $i < j$. As $[x_p, y_q] = [x_p, y_q - \lambda x_p]$ for some λ, one can suppose that $w_1 < w'_1$. In fact w_1 and w'_1 are so-called Lyndon words (see [R]), that is to say satisfy $w_1 w'_1 < w'_1 w_1$. In

$$\begin{aligned} [x_p, y_q] &= \lambda_1 \lambda'_1 (w_1 w'_1 - w'_1 w_1) \\ &+ \lambda_1 \sum_j \lambda'_j (w_1 w'_j - w'_j w_j) + \lambda'_1 \sum_i \lambda_i (w_i w'_1 - w'_1 w_i) + \sum_{i,j>1} \lambda_i \lambda'_j (w_i w'_j - w'_j w_i) \end{aligned} \tag{16}$$

As $w_1 < w'_1 < w'_j$ we deduce that $w_1 w'_1 < w_i w'_1 < w_i w'_j$ for each $i, j > 1$. We have also $w_1 w'_1 < w'_1 w_1 < w'_1 w_i < w'_j w_i$ so $w_1 w'_1$ is the smallest word in formula (16). This proves that $([x_p, y_q], w_1 w'_1) \neq 0$ and so $[x_p, y_q] \neq 0$. □

Remark. — Lemma 4.4 shows that for any not null Lie polynomial P the kernel of ad P is spanned by P.

We will show that

Proposition 4.3 *There exists a sequence of mth-order polygons g_m with even length $l_m \leq m^2$.*

Proof. — We will prove by induction on m the following $P(m)$: "there exists a sequence g_m or order exactly m with even length $l_m \leq m^2$."

If $m = 1$, then $g_1 = x$ or $g_1 = y$ is convenient. If $m = 2$ then $g_2 = (x, y) = xyx^{-1}y^{-1}$ is convenient and has length 4. If $m = 3$ then $g_3 = (g_1, g_2) = x^2 yx^{-1}y^{-1}x^{-1}yxy^{-1} \cdot x^{-1}$ is a third-order polygon so as $xyx^{-1}y^{-1}x^{-1}yxy^{-1}$ that has length 8.

Suppose now $P(m)$.

- If $m + 1 = 2p + 1$ is odd, let us consider $g = (g_p, g_{p+1})$. p and $p + 1$ have not same parity so
$$l(g) \leq 2(l_p + l_{p+1}) \leq 2(p^2 + (p+1)^2 - 1) = (2p+1)^2 - 1.$$
We thus deduce that g is a $(2p+1)$th-order polygon and so $l_{2p+1} \leq (2p+1)^2 - 1$.
- If $m + 1 = 4p$, let us consider $g = (g_{2p-1}, g_{2p+1})$.
$$l(g) \leq 2(l_{2p-1} + l_{2p+1}) \leq 2((2p+1)^2 + (2p-1)^2 - 2) = (4p)^2$$
- If $m + 1 = 4p + 2$, let us consider $g = (g_{2p+1}, \varphi(g_{2p+1}))$. Here φ is the involution $x \mapsto y, y \mapsto x$. If $\mu'(g_{2p+1}) = \exp(\sum_{k \leq 2p+1} x_k)$, we will have
$$\mu'(g) = \exp([x_{2p+1}, \varphi(x_{2p+1})] + \textstyle\sum_{k \geq 2p+2} y_k). \tag{17}$$
The degree of x_{2p+1} in x is not the degree in y so x_{2p+1} as $2p+1$ is odd and $\varphi(x_{2p+1})$ have not same multi-degree thus are not proportional. It follows that $g \in F_{\geq m+1}(X)$. We have
$$l(g) \leq 4 l_{2p+1} \leq (4p+2)^2 - 4 \leq (4p+2)^2. \tag{18}$$

We thus deduce that $l_{m+1} \leq (m+1)^2$. Proposition 4.3 is then proved. □

4.3 Rational series

In fact there is a strong connection with the rank of rational series. The set $\hat{A}(X)$ is usually denoted by $\mathbb{Q}<<X>>$ and is called the set of formal series.

Consider the following operation of X^* on $\hat{A}(X)$; for $u \in X^*$, let

$$u^{-1}S = \sum_{w \in X^*} (S, uw)w \tag{19}$$

We extend it by linearity to obtain $\hat{A}(X)$ as a right module over $A(X)$.

A combinatorial interpretation of that operation in the case where $S = v$ is a single word says that $u^{-1}v$ vanishes , unless v starts with u, that is, $v = uv'$, and in that case $u^{-1}v = v'$.

Definition 4.4 *A formal series is rational if it is an element of the closure of* $A(X)$

A fundamental theorem due to M.-P. Schützenberger assures that the orbits of the action of $A(X)$ are finite dimensional over $\mathbb{Q}$ on rational series. We may then state the following

Definition 4.5 *The rank of a rational series* S *is the dimension of the space* $S \circ A(X)$.

We state now corollary 3.6 of [BR].

Proposition 4.4 *If* $S \in 1 + \hat{A}_{\geq m}(X)$ *is a rational series, then* $\operatorname{rank} S \geq m$

To obtain a lower bound on the length of a polygon we will compute the rank of the rational series $\mu(g) = (1+x)^{a_1}(1+y)^{b_1} \cdots (1+x)^{a_n}(1+y)^{b_n}$.

Proposition 4.5 $\operatorname{rank}\left[(1+x)^{a_1}(1+y)^{b_1} \cdots (1+x)^{a_n}(1+y)^{b_n}\right] \leq \sum_i |a_i| + |b_i|$.

Proof. — We first observe that the following properties are easily established [BR]

$$x^{-1}(ST) = (x^{-1}S)T + (S,1)(x^{-1}T) \tag{20}$$

$$x^{-1}(S^*) = x^{-1}S^* \quad \text{where} \quad S^* = (1-S)^{-1} \tag{21}$$

Observe that $x^{-1}(1+x) = 1, x^{-1}(1+y) = 0, y^{-1}(1+x) = 0, y^{-1}(1+y) = 1$.

An easy computation then gives that $\operatorname{rank}\left[(1+x)^a\right] = |a|$, and this implies that

$$\operatorname{rank}\left[(1+x)^a(1+y)^b\right] = |a| + |b| .$$

From equation 20 we deduce that $\operatorname{rank}(ST) \leq \operatorname{rank}(S) + \operatorname{rank}(T)$ and that implies that the rank of a product $(1+x)^{a_1}(1+y)^{b_1} \cdots (1+x)^{a_n}(1+y)^{b_n}$ can be at most $\sum_{i=1}^n |a_i| + |b_i|$. □

References

[BeR] A. Bellaïche, J.-J. Risler (editors), Sub-Riemannian Geometry. Progress in Mathematics 144. Birkhäuser 1996.

[BR] J. Berstel, C. Reutenauer, Rational series and their languages EATCS Monographs on Theoretical Computer Science. Springer (1988).

[B] N. BOURBAKI, *Groupes et algèbres de Lie*. Hermann, Paris (1972).

[FGR] E. FALBEL, C. GORODSKI, M. RUMIN, *Holonomy of sub-Riemannian manifolds*. International Journal of Mathematics, **8**, No. 3, pp. 317-344, 1997.

[K] P.-V. KOSELEFF, *About approximations of exponentials*, Lecture Notes in Computer Science, **948** pp. 323-333, 1995.

[La] J. P. LABUTE, *Groups and Lie algebras: the Magnus theory* in mathematical legacy of Wilhelm Magnus: groups, geometry and special functions, Contemp. Math. **169**, A.M.S., 1992.

[LS] R.C. LYNDON R.C., P.E. SHUPP, *Combinatorial group theory*, Springer (1977).

[R] C. REUTENAUER, *Free Lie algebras*, Oxford Science Publications (1993).

[Sa] I. N. SANOV, *A property of a representation of a free group.* Dokl. Akad. Nauk. SSSR **57**, 657-659, 1947.

[Su] M. SUZUKI, *General nonsymetric higher-order decompositions of exponential operators and symplectic integrators*, Physics Letters A, **165**, pp. 387-395, 1993

Rigidity and Flexibility of triangle groups in complex hyperbolic geometry

E. FALBEL, P.-V KOSELEFF
Topology 41(4) (2002), pp. 767-786

PERGAMON

Topology 41 (2002) 767–786

TOPOLOGY

www.elsevier.com/locate/top

Rigidity and flexibility of triangle groups in complex hyperbolic geometry

E. Falbel*, P.-V. Koseleff

Institut de Mathématiques, Analyse Algebrique, Université Pierre et Marie Curie, 4, place Jussieu, 175 rue du Chevaleret, F-75252 Paris, France

Received 13 December 1999; received in revised form 30 June 2000; accepted 5 September 2000

Abstract

We show that the Teichmüller space of the triangle groups of type (p, q, ∞) in the automorphism group of the two-dimensional complex hyperbolic space contains open sets of 0, 1 and two real dimensions. In particular, we identify the Teichmüller space near embeddings of the modular group preserving a complex geodesic.

MSC: 32H20; 20H10; 22E40; 57S30; 32G07; 32C16

Keywords: Complex hyperbolic; Triangle group; Discrete group; Rigidity; CR-manifolds

1. Introduction

Let Γ be the triangular group of type (p, q, ∞), that is, the abstract group presented by

$$\langle \iota_0,\iota_1,\iota_2 \colon \iota_0^2 = 1, \iota_1^2 = 1, \iota_2^2 = 1, (\iota_0 \circ \iota_1)^p = 1, (\iota_0 \circ \iota_2)^q = 1 \rangle.$$

The faithful discrete embeddings of Γ in the isometry group of the one-dimensional complex disc, that is $\widehat{\mathbf{PU}(1,1)}$ (containing the holomorphic and the anti-holomorphic transformations), with anti-holomorphic embedding of generators and $\iota_1 \circ \iota_2$ embedded as a parabolic, are rigid.

We consider in this paper faithful discrete embeddings of Γ in $\widehat{\mathbf{PU}(2,1)}$ (containing the holomorphic and the anti-holomorphic transformations), the isometry group of the two-dimensional

* Corresponding author. Fax: + 33-1-44-27-85-51.
E-mail addresses: falbel@math.jussieu.fr (E. Falbel), koseleff@math.jussieu.fr (P.-V. Koseleff).

PII: S0040-9383(00)00049-5

complex ball (with the natural metric invariant under biholomorphisms). Let the *Teichmüller space* denote the space of faithful discrete embeddings modulo conjugation in $\widehat{\mathbf{PU}(2,1)}$, with anti-holomorphic generators and such that $\iota_1 \circ \iota_2$ is parabolic. We obtain the following description.

Theorem 1.1. *The Teichmüller space of the triangle groups of type* (p, q, ∞), *with* $2 < p \leqslant q$, *in the automorphism group of the two-dimensional complex hyperbolic space contains open sets of* 0, 1 *and* 2 *real dimensions*.

Each open set of Teichmüller space in the theorem contains a **C**-Fuchsian embedding, that is, an embedding which fixes a complex geodesic setwise.

The triangle groups of type $(2, p, \infty)$ are special. The involution of order 2 cannot be deformed and we loose one parameter in the deformation space. Observing that the index 2 subgroup of holomorphic transformations of the triangle group of type $(2, 3, \infty)$ is the modular group $\mathbf{SL}(2, \mathbf{Z})$, we obtain the following:

Theorem 1.2. *The Teichmüller space of the modular group in the biholomorphic automorphism group of the two-dimensional complex hyperbolic space around a representation that fixes a complex geodesic is of dimension* 0 *or* 1.

Of course, we impose the parabolic generator to be represented by parabolics. We will describe explicitly the embeddings in each case of the theorem.

It is important to contrast that result with the rigidity result of [3,5,12]. If the group has a **C**-Fuchsian embedding which is cocompact in the fixed complex geodesic, any nearby deformation is **C**-Fuchsian. In our case, the volume is finite. It is interesting that both rigidity and flexibility occur and depend on the particular **C**-Fuchsian embedding.

Embeddings of triangle groups of type (∞, ∞, ∞) were previously analyzed in [6,7] and [1]. It was shown in [1] that the Teichmüller space, in that case, contains an open set of real dimension 4. Deformations of the modular group were also obtained by Parker [10] independently. The difficulty in the present case is the appearance of elliptic transformations. The proof is based on a Poincaré's polyhedron theorem for complex hyperbolic geometry developed in [2]. We construct explicitly the fundamental domains. The idea behind it is that anti-holomorphic reflections, which fix real geodesics are the analog of inversions in classical conformal geometry. Polyhedra are constructed having faces, foliated by complex geodesics, invariant under those reflections. They are a generalization of Mostow's bisectors and we will call them **C**-spheres. Edges are complex geodesics in the intersection of two **C**-spheres.

We hope that the methods of this paper will achieve (see Remark 5.5, p. 17), in the future, a precise description of Teichmüller space in the case of embeddings by anti-holomorphic transformations. On the other hand, it would be interesting to obtain our results using a modification of the method of Higgs bundles (see [8]) for non-compact surfaces.

An interesting problem would be to understand the behavior of automorphic forms under the deformation of modular group.

2. The complex hyperbolic space and its boundary

In this section and the following, we collect general results about the complex hyperbolic space. As a reference we use [4,9,1].

2.1. **PU**(2,1), *the Heisenberg group and the Cayley transform*

Let $\mathbf{C}^{n+2}$ denote the complex vector space equipped with the Hermitian form

$$b(z,w) = -\bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_{n+2} w_{n+2}.$$

Consider the following subspaces in $\mathbf{C}^{n+2}$:

$$\mathrm{V}_0 = \{z \in \mathbf{C}^{n+2}\colon b(z,z) = 0\},$$

$$V = \{z \in \mathbf{C}^{n+2}\colon b(z,z) < 0\}.$$

Let $P\colon \mathbf{C}^{n+2}\backslash\{0\} \to \mathbf{C}P^{n+1}$ be the canonical projection onto the complex projective space. Then $H_{\mathbf{C}}^{n+1} = P(V)$ equipped with the Bergman metric is the complex hyperbolic space. The orientation preserving isometry group of $H_{\mathbf{C}}^{n+1}$ is generated by $\mathbf{PU}(n+1,1)$, the unitary group of b and the anti-holomorphic transformations. We denote it by $\widehat{\mathbf{PU}(n,1)}$. Also, $\mathbf{PU}(n+1,1)$ is the group of biholomorphic transformations of $H_{\mathbf{C}}^{n+1}$. Let $S^{2n+1} = P(V_0)$. Then S^{2n+1} is the boundary of $H_{\mathbf{C}}^{n+1}$. We may consider $H_{\mathbf{C}}^{n+1}$ and S^{2n+1} as the unit ball and the unit sphere in $\mathbf{C}^{n+1}$.

We restrict our attention to the two-dimensional complex case and in the following we use the conventions of [9] (see also [4]). The mapping

$$C\colon (w_1, w_2) \mapsto \left(z_1 = \frac{\mathrm{i}w_1}{1+w_2}, z_2 = \mathrm{i}\frac{1-w_2}{1+w_2} \right)$$

is usually referred to as the *Cayley transform*. It maps the unit ball

$$B = \{w \in \mathbf{C}^2\colon |w_1|^2 + |w_2|^2 < 1\}$$

biholomorphically onto

$$V = \{z \in \mathbf{C}^2\colon \mathrm{Im}(z_2) > |z_1|^2\}.$$

The Cayley transform leads to a generalized form of the *stereographic projection*. This mapping $\pi\colon S^3\backslash\{-e_2\} \to \mathbf{R}^3$, where $S^3 = \partial B$ and $e_2 = (0,1) \in \mathbf{C}^2$, is defined as the composition of the Cayley transform restricted to $S^3\backslash\{-e_2\}$ followed by the projection:

$$(z_1, z_2) \mapsto (z_1, \mathrm{Re}(z_2)).$$

The stereographic projection π can be extended to a mapping from S^3 onto the one-point compactification $\bar{\mathbf{R}}^3$ of $\mathbf{R}^3$.

The *Heisenberg group* $\mathbf{H}$ is the set of pairs $(z,t) \in \mathbf{C} \times \mathbf{R}$ with the product

$$(z,t)\cdot(z',t') = (z+z', t+t' + 2\,\mathrm{Im}\, z\bar{z}').$$

Using the stereographic projection, we can identify $S^3\backslash\{-e_2\}$ with $\mathbf{H}$ and S^3 with the one-point compactification $\bar{\mathbf{H}}$ of $\mathbf{H}$. The inverse function of the stereographic projection is given by

$$\pi^{-1}(z,t)=\left(\frac{-2\mathrm{i}z}{1+|z|^2-\mathrm{i}t},\frac{1-|z|^2+\mathrm{i}t}{1+|z|^2-\mathrm{i}t}\right).$$

Observe that the x-axis in the Heisenberg group corresponds to the intersection of S^3 with the real plane $\mathrm{Re}(w_1)=0$, $\mathrm{Im}(w_2)=0$. Also, the y-axis corresponds to the intersection of S^3 with the real plane $\mathrm{Im}(w_1)=0$, $\mathrm{Im}(w_2)=0$.

The Heisenberg group acts on itself by left translations. Heisenberg translations by $(0,t)$ for $t\in\mathbf{R}$ are called *vertical translations*.

Positive scalars $\lambda\in\mathbf{R}_+$ act on $\mathbf{H}$ by *Heisenberg dilations*

$$d_\lambda:(z,t)\mapsto(\lambda z,\lambda^2 t).$$

If $\rho\in\mathbf{U}(1)$, then ρ acts on $\mathbf{H}$ by *Heisenberg rotation*

$$\rho:(z,t)\mapsto(\rho z,t).$$

The *Heisenberg complex inversion* of $\mathbf{H}$ is defined on $\mathbf{H}\backslash\{(0,0)\}$ by

$$h:(z,t)\mapsto\left(\frac{-z}{|z|^2-\mathrm{i}t},-\frac{t}{|z|^4+t^2}\right).$$

Note that $h=\pi\circ j\circ\pi^{-1}$, where j is the involution

$$j:(w_1,w_2)\mapsto(-w_1,-w_2),\quad (w_1,w_2)\in\mathbf{C}^2.$$

The map $\hat{m}$ defined by

$$\hat{m}:(z,t)\mapsto(\bar{z},-t),$$

corresponds to

$$\pi^{-1}\circ\hat{m}\circ\pi(w_1,w_2)=(-\bar{w}_1,\bar{w}_2).$$

All these actions extend trivially to the compactification $\bar{\mathbf{H}}$ of $\mathbf{H}$. It is well known that the group G of transformations of $\bar{\mathbf{H}}$ generated by all Heisenberg translations, dilations, rotations, and h coincides with $\pi^{-1}\circ\mathbf{PU}(2{,}1)\circ\pi$, and the group $\hat{G}=\langle G,\hat{m}\rangle$ is the group of all conformal transformations of $\bar{\mathbf{H}}$ (see [9,4]).

We need explicit representations of the matrices corresponding to transformations on $\mathbf{SU}(2{,}1)$.

The transformations $R_\theta:(z,t)\mapsto(\exp(\mathrm{i}\theta)z,t)$ and $T_{z,t}:(z',t')\mapsto(z,t).(z',t')$ are represented, respectively, by the following matrices in $\mathbf{SU}(2{,}1)$.

$$R_\theta=\begin{pmatrix}\exp(-\mathrm{i}\theta/3)&0&0\\0&\exp(2\mathrm{i}\theta/3)&0\\0&0&\exp(-\mathrm{i}\theta/3)\end{pmatrix},$$

$$T_{z,t}=\begin{pmatrix}1+|z|^2/2-\mathrm{i}t/2&\mathrm{i}\bar{z}&|z|^2/2-\mathrm{i}t/2\\-\mathrm{i}z&1&-\mathrm{i}z\\-|z|^2/2+\mathrm{i}t/2&-\mathrm{i}\bar{z}&1-|z|^2/2+\mathrm{i}t/2\end{pmatrix}.$$

The anti-holomorphic transformations on the ball $(z_1, z_2) \mapsto (\bar{z}_1, -\bar{z}_2)$ (which corresponds to the standard inversion, see the next section) and $\pi^{-1} \circ \hat{m} \circ \pi : (z_1, z_2) \mapsto (-\bar{z}_1, \bar{z}_2)$ correspond, respectively, to the matrices in $\mathbf{SU}(2,1)$

$$I_0 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad I_x = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Their action on homogeneous coordinates should be preceded by the conjugation.

We will use the following proposition that characterizes certain elliptic elements of $\mathbf{PU}(2,1)$. We say that a matrix in $\mathbf{SU}(2,1)$ is elliptic if it is conjugate to an element of $\mathbf{U}(2)$ (see [4]).

Proposition 2.1 (Goldman [4]). *Let $A_1, A_2 \in \mathbf{SU}(2,1)$ be elliptic elements. Then they are conjugate if and only if* $\mathrm{tr}(A_1) = \mathrm{tr}(A_2)$.

Observe that this implies that they are conjugate in $\mathbf{PU}(2,1)$ if and only if the cube of their traces are equal.

3. R-circles, C-circles and C-spheres

There are two kinds of totally geodesic submanifolds of real dimension 2 in $H^2_{\mathbf{C}}$: *complex geodesics* (represented by $H^1_{\mathbf{C}} \subset H^2_{\mathbf{C}}$) and *totally real geodesic* 2*-planes* (represented by $H^2_{\mathbf{R}} \subset H^2_{\mathbf{C}}$). Each of these totally geodesic submanifold is a model of the hyperbolic plane.

Consider the complex hyperbolic space $H^2_{\mathbf{C}}$ and its boundary $\partial H^2_{\mathbf{C}} = S^3$. We will call **C**-*circles* the intersections of S^3 with the boundaries of totally geodesic complex submanifolds $H^1_{\mathbf{C}}$ in $H^2_{\mathbf{C}}$. Analogously, we will call **R**-*circles* the intersections of S^3 with the boundaries of totally geodesic totally real submanifolds $H^2_{\mathbf{R}}$ in $H^2_{\mathbf{C}}$.

3.1. **R**-*circles*

Definition 3.1. An inversion on an **R**-circle is a non-trivial conformal transformation which fixes it pointwise.

Observe that an inversion has invariant **R**-circles, one of them being pointwise fixed. Moreover, an **R**-circle defines a unique inversion. There is, then, a one-to-one correspondence between inversions and **R**-circles. For instance, the transformation $\hat{m}(z, t) = (\bar{z}, -t)$ on the Heisenberg group is the inversion that fixes pointwise the **R**-circle $\mathrm{Im}(z) = 0$.

Proposition 3.2 (Falbel and Zocca [2]). *Let I_1 and I_2 be reflections on the* **R**-*circles* $\mathbf{R}_1$ *and* $\mathbf{R}_2$,

(i) *$I_1 \circ I_2$ is parabolic if and only if $\mathbf{R}_1$ and $\mathbf{R}_2$ intersect at one point.*
(ii) *$I_1 \circ I_2$ is loxodromic if and only if $\mathbf{R}_1$ and $\mathbf{R}_2$ do not intersect and are not linked.*
(iii) *$I_1 \circ I_2$ is elliptic if and only if $\mathbf{R}_1$ and $\mathbf{R}_2$ are linked or intersect at two points. In the first case there are two* (*exactly two in most cases*) **C**-*circles setwise invariant under both inversions. In the last case there exists one pointwise invariant* **C**-*circle under $I_1 \circ I_2$.*

Proof. See [2] for the proof. Observe that an elliptic element is conjugated to an element of $\mathbf{U}(2)$ which generically has two one-dimensional complex eigenspaces where the action is just a rotation. □

Definition 3.3. Let I_1 and I_2 be reflections on the **R**-circles $\mathbf{R}_1$ and $\mathbf{R}_2$ linked or intersecting twice. We say that $I_1 \circ I_2$ has type (φ_1, φ_2) if it is a rotation of angles φ_1 and φ_2 on the invariant **C**-circles.

3.2. ***C**-circles and **C**-spheres*

Proposition 3.4 (see Goldman [4]). *In the Heisenberg model*, **C***-circles are either vertical lines or ellipses*, *whose projection on the z-plane are circles.*

Definition 3.5. The contact plane at $M = (a, b, c)$ is the plane $P(M) := Z - t + 2aY - 2bX$.

The circle of *center* $M = (a, b, c)$ and *radius* R is the intersection of the contact plane at M and the cylinder $(X - a)^2 + (Y - b)^2 = R^2$.

Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two circles of centers (a_1, b_1, c_1), (a_2, b_2, c_2) and *radii* R_1 and R_2. Let d and h be the horizontal and vertical distances between centers

$$d = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}, \quad h = c_2 - c_1, \quad S = \tfrac{1}{2}(a_1 b_2 - a_2 b_1).$$

Proposition 3.6 (Linking of **C**-circles, Falbel and Koseleff [1]). *The* **C***-circles* $\mathbf{C}_1$ *and* $\mathbf{C}_2$ *are linked if and only if*

$$\begin{aligned}(d^2 - (R_1 + R_2)^2)(d^2 - (R_1 - R_2)^2) + (h + 4S)^2 \\ = (d^2 - (R_2^2 - R_1^2))^2 + (h + 4S)^2 - 4d^2 R_1^2 < 0.\end{aligned}$$

Observe that $\mathbf{C}_1$ and $\mathbf{C}_2$ are not linked if their projections are not, that is

$$(d^2 - (R_1 + R_2)^2)(d^2 - (R_1 - R_2)^2) > 0$$

or if

$$4d^2 R_1^2 < (h + 4S)^2,$$

that is, $\mathbf{C}_1$ does not intersect the plane defining $\mathbf{C}_2$ (see also Lemma 6.1).

In the following definition we allow a point to be a (degenerate) **C**-circle.

Definition 3.7. A **C**-sphere around an **R**-circle is an union of **C**-circles invariant under the inversion on the **R**-circle, which is homeomorphic to a sphere. We will call the axis of the **C**-sphere the set of centers of these invariant **C**-circles.

In particular, a **C**-sphere contains two degenerate **C**-circles and its axis has starting and ending points in the **R**-circle. See also [2].

Definition 3.8. The surface of centers of an **R**-circle is the set of points which are the centers of invariant **C**-circles under the inversion on the **R**-circle.

Such **C**-circles have two points in common with the **R**-circle. Observe that for finite **R**-circles this is a two-dimensional surface but for an infinite **R**-circle this coincides with the **R**-circle.

For a finite **R**-circle the center completely determines the **C**-circle (Proposition 3.9). But, observe that for an infinite **R**-circle, a *radius* should be specified for each center.

A given axis determines a surface obtained by the union of **C**-circles defined by the centers. But that surface might have self-intersections. We will call a *good axis* an axis whose associated surface is homeomorphic to the two-dimensional sphere.

We will consider parts of **C**-spheres as faces of polyhedra. By abuse of language we will also refer to them as **C**-spheres. There will be a disjoint union of **C**-circles between two fixed **C**-circles in a **C**-sphere. Analogously, we will refer to the part of the axis corresponding to that portion of **C**-sphere as the corresponding axis.

3.3. *Standard* **R**-*circle*

Consider the following transformation on the Heisenberg group:

$$I_0 = \hat{m} \circ h : (z,t) \mapsto \left(\frac{-\bar{z}}{|z|^2 + \mathrm{i}t}, \frac{t}{|z|^4 + t^2} \right).$$

which corresponds to

$$\pi^{-1} \circ I \circ \pi(w_1, w_2) = \pi^{-1} \circ \hat{m} \circ h \circ \pi(w_1, w_2) = (\bar{w}_1, -\bar{w}_2).$$

I_0 leaves pointwise fixed the standard **R**-circle $\mathbf{R}_0$ (see [4] for details) (Fig. 1)

$$r^2 + \mathrm{i}t = -\mathrm{e}^{-2\mathrm{i}\theta},$$

where $z = r\mathrm{e}^{\mathrm{i}\theta}$. In cylindrical coordinates $\mathbf{R}_0$ is given by

$$r = \sqrt{-\cos 2\theta}, \quad z = \sin 2\theta.$$

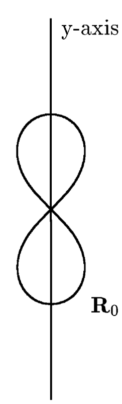


Fig. 1. $\mathbf{R}_0$ top view.

Proposition 3.9 (Falbel and Koseleff [1]). *The surface of centers* $\mathbf{S}_0$ *of the standard* $\mathbf{R}$*-circle is given, in cylindrical coordinates, by* $(r,\theta,\sin 2\theta)$. *If* $r^2+\cos 2\theta \geqslant 0$, *the radius of a* $\mathbf{C}$*-circle with center coordinates* r,θ *is* $\sqrt{r^2+\cos 2\theta}$.

From the above parametrization we see that the surface of centers satisfies the algebraic equation

$$Z(X^2+Y^2)=2XY.$$

The *radius* of the $\mathbf{C}$-circle is then $R^2=r^2+\cos 2\theta$. We obtain the standard $\mathbf{R}$-circle for $R=0$.

The part of the algebraic set (solution of the equation above) with $r^2+\cos 2\theta<0$ will be called *surface of imaginary centers*. Although it does not correspond to any center of an invariant $\mathbf{C}$-circle, it is a useful set.

Schwarz (cf. [11]) calls hybrid cones certain surfaces foliated by $\mathbf{R}$-circles. Part of our surface of centers is a hybrid cone in his sense. We make use of them in this work mainly to parametrize $\mathbf{C}$-spheres, while his use of them is as boundaries of fundamental domains. One could imagine that probably a complete description of fundamental domains in complex hyperbolic geometry should take into account both $\mathbf{C}$-spheres, foliated by $\mathbf{C}$-circles, and surfaces foliated by $\mathbf{R}$-circles, as Schwarz's hybrid cones.

3.4. *Infinite* $\mathbf{R}$*-circles*

Definition 3.10. $\mathbf{R}_{r,\theta,\alpha}$ is the infinite $\mathbf{R}$-circle passing through $M=(r,\theta,\sin 2\theta)$ whose projection onto the z-plane is the line of slope $\tan\alpha$. It is given by

$$P(M)\colon Z=\sin 2\theta+2r\cos\theta Y-\sin\theta X, \quad P_\alpha(M)\colon \sin\alpha X-\cos\alpha Y=r\sin(\theta-\alpha).$$

Observe that $\mathbf{R}_{r,\theta,\alpha}$ is horizontal if and only if $\alpha=\theta \bmod \pi$.

Proposition 3.11 (Linking of $\mathbf{R}$-circles). *If* $R^2=r^2+\cos 2\theta<0$, *then* $\mathbf{R}_{r,\theta,\alpha}$ *is horizontal and intersects* $\mathbf{R}_0$ *twice or is linked with* $\mathbf{R}_0$. *On the other hand, any* $\mathbf{R}$ *-circle that is linked with* $\mathbf{R}_0$ *intersects the surface of imaginary centers once.*

Proof. Let $\mathbf{R}=\mathbf{R}_{r,\theta,\alpha}$ and $M=(r,\theta,\sin 2\theta)$. Let us consider the intersection of $\mathbf{R}_0$ with $P(M)$. We obtain points $(\rho=\sqrt{-\cos 2u},u,\sin 2u)$, such that $\sin(\theta-u)(\cos(\theta+u)+r\rho)=0$. We get the two points $(\pm\sqrt{-\cos 2\theta},\theta,\sin 2\theta)$ or $(\cos(\theta+u)+r\rho)=0$, that gives $(r^2+\cos 2\theta)\cos 2u+\sin(\theta-u)^2=0$, which is impossible. Now the plane $P_\alpha(M)$ separates the two points if $\alpha\neq\theta[\pi]$. $\square$

3.5. *Infinite* $\mathbf{R}$*-circles and surface of centers*

Let $\mathbf{R}$ be an infinite $\mathbf{R}$-circle passing through $\mathbf{S}_0$ at $M_1=(r_1,\theta_1,\sin 2\theta_1)$ and $M_2=(r_2,\theta_2,\sin 2\theta_2)$ (see Fig. 2). Then we have

$$\sin(\theta_1-\theta_2)(\cos(\theta_1+\theta_2)+r_1r_2)=0.$$

If $\sin(\theta_1-\theta_2)=0$, then $\mathbf{R}$ is horizontal and intersects $\mathbf{S}_0$ at $r=0$ again.

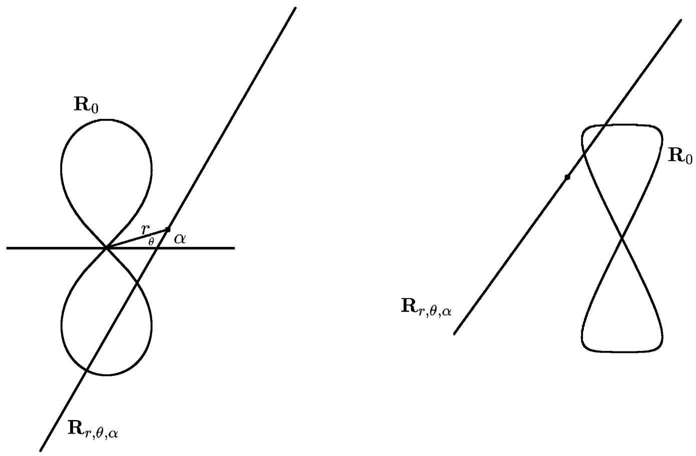


Fig. 2. $\mathbf{R}_{r,\theta,\alpha}$; lateral view (right) and top view (left). The distinguished point is an intersection of the infinite **R**-circle with the surface of centers of $\mathbf{R}_0$.

Proposition 3.12. *Let* **R** *be an infinite* **R**-*circle intersecting* $\mathbf{S}_0$ *twice. Then* **R** *intersects* $\mathbf{S}_0$ *in the third point and we have*

$$\frac{1}{R_1^2}+\frac{1}{R_2^2}+\frac{1}{R_3^2}=0, \quad \frac{\sin 2\theta_1}{R_1^2}+\frac{\sin 2\theta_2}{R_2^2}+\frac{\sin 2\theta_3}{R_3^2}=0,$$

where $M_i=(r_i,\theta_i,\sin 2\theta_i)$ *are the intersection points and* $R_i^2=r_i^2+\cos 2\theta_i$.

Proof. We consider the polynomial equations

$$Z(X^2+Y^2)=2XY, \quad r^2=(X^2+Y^2), \quad TX=Y,$$

$$Z=\sin 2\theta_1+2r_1(\cos\theta_1 Y-\sin\theta_1 X),$$

$$Z=\sin 2\theta_2+2r_2(\cos\theta_2 Y-\sin\theta_2 X).$$

Using different elimination techniques, we get a third-degree polynomial equation that is satisfied by r^2, a third-degree polynomial equation for Z and a third-degree polynomial equation for T. As we know already two roots of these polynomials, we get (using the fact that $r_1r_2+\cos(\theta_1+\theta_2)=0$)

$$Z_3=\frac{R_1^2\sin 2\theta_2+R_2^2\sin 2\theta_1}{R_1^2+R_2^2}, \quad r_3^2=\frac{\sin(\theta_1-\theta_2)^2}{R_1^2+R_2^2}, \quad T_3=\frac{r_1\cos\theta_2-r_2\cos\theta_1}{r_1\sin\theta_2-r_2\sin\theta_1}.$$

We then obtain $R_3^2=r_3^2+(1-T_3^2)/(1+T_3^2)=-R_1^2R_2^2/(R_1^2+R_2^2)$. □

Observe that in this case one of the intersection points is an imaginary center and that **R** is linked with $\mathbf{R}_0$ (Proposition 3.11).

On the other hand, we get (see Proposition 3.2).

Proposition 3.13. *Let* $\mathbf{R}$ *be an infinite non-horizontal* $\mathbf{R}$*-circle, linked with* $\mathbf{R}_0$*. Then* $\mathbf{R}$ *intersects the surface of centers at three points. One is imaginary and the other corresponding* $\mathbf{C}$*-circles are linked.*

Proof. Let $\mathbf{R} = \mathbf{R}_{r,\theta,\alpha}$. We have $R^2 = r^2 + \cos 2\theta < 0$. Points of $\mathbf{R}$ are

$$X = r\cos\theta - \frac{1}{2}\frac{\cos\alpha(Z - \sin 2\theta)}{r\sin(\alpha - \theta)}, \quad Y = r\sin\theta - \frac{1}{2}\frac{\sin\alpha(Z - \sin 2\theta)}{r\sin(\alpha - \theta)}.$$

Looking to the intersection of $\mathbf{R}$ and $Z(X^2 + Y^2) = 2XY$, one finds a polynomial P of degree 2 whose discriminant is

$$16\sin(\alpha - \theta)^2(-4r^4\sin(\alpha - \theta)^2 + \cos(\alpha + \theta)^2 + 4r^2\sin(-\alpha + \theta)\sin(\alpha + \theta))$$

and is positive when $0 \leqslant r^2 < -\cos 2\theta$. We thus obtain two other points M_1 and M_2.

Looking to the corresponding *radii* of $\mathbf{C}$-circles we get $P = (R^2 - R_1^2)(R^2 - R_2^2)$ with $R_1^2 R_2^2 > 0$ when $0 \leqslant r^2 < -\cos 2\theta$. It shows that R_1^2 and R_2^2 are positive because of Proposition 3.12 and $R_3^2 = r^2 + \cos 2\theta < 0$.

The evaluation of the linking condition (Proposition 3.6) gives $-2(r_1^2 + \cos 2\theta_1)(r_2^2 + \cos 2\theta_2) < 0$ so the $\mathbf{C}$-circles at M_1 and M_2 are linked. □

4. Configurations of a standard and an infinite R-circle

Consider the configuration space of an infinite $\mathbf{R}$-circle $\mathbf{R}_1$ and the standard $\mathbf{R}$-circle $\mathbf{R}_0$. Using a description of the configuration space of lines in the plane as a Möbius band (take as coordinates the angle, between 0 and π, from the x-axis and the oriented measure of a segment starting from the origin and arriving perpendicularly to a line) one can clearly obtain, by vertical translations, that the configuration space is $M \times \mathbf{R}$, where M is the Möbius band. One could further use the dihedral symmetry of the standard $\mathbf{R}$-circle (generated by reflections on the the two horizontal x- and y-axis) to reduce the configuration space to angles between 0 and $\pi/2$ and segments of positive measures.

Observe that the configuration space above is not the configuration space of two $\mathbf{R}$-circles up to transformations of $\mathbf{PU}(2,1)$. As one of the $\mathbf{R}$-circles is infinite, the transformations are in the isotropy group $\mathbf{U}(1) \rtimes \mathrm{H}$.

We will deal only with configurations that give elliptic elements. By Proposition 3.2, the infinite $\mathbf{R}$-circles should be linked with the standard $\mathbf{R}$-circle.

It will be important to identify equivalent configurations under $\mathbf{PU}(2,1)$. Before doing that, we single out some special configurations which will represent each equivalent class (see Fig. 3).

Definition 4.1. The $\mathbf{C}$-standard elliptic (φ_1,φ_2)-configuration is the one where $\mathbf{R}_1 = \mathbf{R}_{r,\theta,\alpha}$ with $r = 0$, $\theta = \pi/4 - \varphi_1/2$, $\alpha = \pi/4 + \varphi_1/2 - \varphi_2$.

Proposition 4.2. *The composition* $I_0 \circ I_{(\varphi_1,\varphi_2)}$ *of the two inversions corresponding to* $\mathbf{R}_0$ *and the* $\mathbf{C}$*-standard elliptic* (φ_1,φ_2)*-configuration* R_1 *is of type* $(2\varphi_1,2\varphi_2)$ (*see Definition* 3.3).

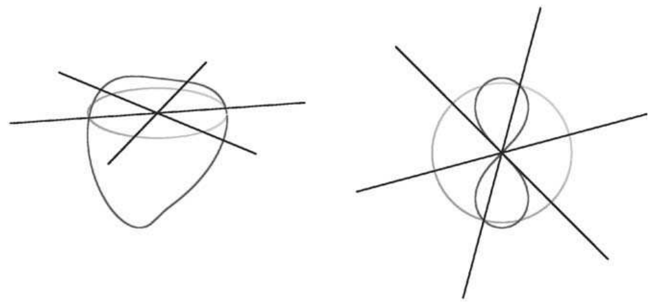


Fig. 3. **C**-standard configurations $(\pi/3, n\pi/3)$, $0 \leqslant n \leqslant 2$; two views showing the invariant **C**-circle. The configurations $(\pi/3,0)$ whose line intersects twice the finite **R**-circle and $(\pi/3,\pi/3)$ are rigid.

In particular, the **C**-standard $(\varphi_1,0)$-configuration corresponds to the horizontal line at height $\sin(\pi/2 - \varphi_1)$ intersecting the standard **R**-circle twice.

The proof of this proposition follows from the following elementary lemma.

Lemma 4.3. *The angle of tangency to the standard* **R**-*circle* $\mathbf{R}_0$ *at the point with polar coordinate* θ *is* $3\theta - \pi/2$. *The angle from the ray at angle* θ *to the tangent to the* **R**-*circle* $\mathbf{R}_0$ *at the intersection point is* $2\theta - \pi/2$.

We will determine those configurations equivalent to a fixed standard one. We start with a rigidity result.

Theorem 4.4 (Rigidity). *The standard* $(\varphi_1,0)$-*configuration* (*for* $\pi/4 \leqslant \varphi_1 \leqslant \pi/2$) *is the unique configuration corresponding to its conjugacy class.*

Proof. Observe that, if the conjugacy class of the composition of inversions is determined by the angles φ_1 and 0, the two **R**-circles intersect. A simple computation then shows that the only element in the conjugacy class is the one defined by the standard one. □

To study more general conjugacy classes we consider the subgroup generated by two inversions that is, I_0 and $M \circ I_x \circ \bar{M}^{-1}$, where $M = T \circ D \circ R$ (R is a rotation, D is a dilation and T is a Heisenberg translation). The corresponding matrix in $\mathbf{SU}(2,1)$ is given by $A = M \circ R_x \circ \bar{M}^{-1} \circ R_0$. One computes

$$M = \frac{\exp(-\mathrm{i}\alpha/3)}{\lambda}\begin{pmatrix} \frac{1}{2}(\lambda^2(1+|z|^2-\mathrm{i}t)+1) & \mathrm{i}\lambda\exp(\mathrm{i}\alpha)\bar{z} & \frac{1}{2}(\lambda^2(1+|z|^2-\mathrm{i}t)-1) \\ -\mathrm{i}\lambda^2 z & \lambda\exp(\mathrm{i}\alpha) & -\mathrm{i}\lambda^2 z \\ \frac{1}{2}(\lambda^2(1-|z|^2+\mathrm{i}t)-1) & -\mathrm{i}\lambda\exp(\mathrm{i}\alpha)\bar{z} & \frac{1}{2}(\lambda^2(1-|z|^2+\mathrm{i}t)+1) \end{pmatrix},$$

$$A = \exp(-2/3\mathrm{i}\alpha)\begin{pmatrix} 1+|z|^2-\mathrm{i}t-\exp(2\mathrm{i}\alpha)\bar{z}^2 & \mathrm{i}(z-\exp(2\mathrm{i}\alpha)\bar{z}) & -|z|^2+\mathrm{i}t+\exp(2\mathrm{i}\alpha)\bar{z}^2 \\ -\mathrm{i}(z-\exp(2\mathrm{i}\alpha)\bar{z}) & -\exp(2\mathrm{i}\alpha) & \mathrm{i}(z-\exp(2\mathrm{i}\alpha)\bar{z}) \\ -|z|^2+\mathrm{i}t+\exp(2\mathrm{i}\alpha)\bar{z}^2 & \mathrm{i}(z-\exp(2\mathrm{i}\alpha)\bar{z}) & -1+|z|^2-\mathrm{i}t-\exp(2\mathrm{i}\alpha)\bar{z}^2 \end{pmatrix}.$$

When $z = re^{i\theta}$ and $t = \sin 2\theta$, we get

$$\begin{aligned}\mathrm{tr}(A) &= e^{-(2/3)i\alpha}(-2i\sin 2\theta + 2r^2 - 2r^2 e^{-2i(\theta-\alpha)} - e^{2i\alpha})\\ &= e^{(1/3)i(\alpha-3\theta)}(-4i(r^2+\cos 2\theta)\sin(\alpha-\theta) + e^{i(\alpha-3\theta)} - 2e^{-i(\alpha-3\theta)}).\end{aligned}$$

We thus deduce that

Proposition 4.5. *When* $\alpha_0 - \theta_0 \neq 0$ (*which excludes the situation of Theorem* 4.4 *and the* **C**-*standard* $(\pi/2,\pi/2)$-*configuration*), A and A_0 *are in the same conjugacy class*, *if*

$$\alpha - 3\theta = \alpha_0 - 3\theta_0, \quad r^2 + \cos 2\theta = (r_0^2 + \cos 2\theta_0)\frac{\sin(\alpha_0-\theta_0)}{\sin(\alpha-\theta)},$$

or equivalently

$$\alpha - 3\theta = \alpha_0 - 3\theta_0, \quad r^2 = \frac{1}{2}\frac{\sin(\alpha_0+\theta_0) - \sin(\alpha+\theta)}{\sin(\alpha-\theta)} + r_0^2\frac{\sin(\alpha_0-\theta_0)}{\sin(\alpha-\theta)}.$$

Observe that in this case we have $\mathrm{tr}(A)^3 = \mathrm{tr}(A_0)^3$. We also have other solutions considering that an infinite **R**-circle $R_{r,\theta,\alpha}$ may be given by $\alpha \bmod \pi$ and by three values of (r,θ).

We consider now the family $\mathbf{R}_{r,\theta,\alpha}$ of infinite **R**-circles defined by

$$\alpha = 3\theta + (\alpha_0 - 3\theta_0) = 3\theta + \pi/4 + \varphi_1/2 - \varphi_2 - 3(\pi/4 - \varphi_1/2) = 3\theta - \pi/2 + 2\varphi_1 - \varphi_2$$

and

$$r^2 = \frac{1}{2}\frac{\sin(\alpha_0+\theta_0) - \sin(\alpha+\theta)}{\sin(\alpha-\theta)} = -\frac{\cos(2\theta+\varphi_1-\varphi_2)\cos(2\theta+\varphi_1)}{\cos(2\theta+2\varphi_1-\varphi_2)}.$$

Remark 4.6. The curve defined above has three branches (the three intersection points with the surface of centers). The intersection with the surface of imaginary centers is a closed curve. Each component has $\alpha - 3\theta$ as invariant. As the sign of r^2 changes at $\theta = \theta_0$ if $r_0 = 0$ one can take $r = r^2/\sqrt{|r^2|}$ as continuous parametrization.

Observe that when $\theta = \pi/4 - \varphi_1/2$ we obtain precisely the **C**-standard (φ_1,φ_2)-configuration.

Theorem 4.7 (Rigidity). *The* **C**-*standard* (φ,φ)-*configuration is rigid in its conjugacy class.*

Proof. The proof follows from solving the equation $\mathrm{tr}(A) = -1$. From the solution above, we obtain that, if $\varphi_1 = \varphi_2$, $r^2 = -\cos\theta$. That is, the other solution gives a parabolic configuration. □

Theorem 4.8 (Flexibility). *The family of compositions* $I_0 \circ I_{\varphi_1,\varphi_2}(\theta)$, *where* $I_{\varphi_1,\varphi_2}(\theta)$ *denotes the inversion on the* **R**-*circle in the family above*, *parametrized by* θ, *is in the conjugacy class defined by the angles* $2\varphi_1$ and $2\varphi_2$.

Proof. Using the Proposition 4.5, we have in **SU**(2,1) $\mathrm{tr}(I_0 \circ \widetilde{\mathrm{I}_{\varphi_1,\varphi_2}}(\theta)) = \mathrm{tr}((I_0 \circ \widetilde{\mathrm{I}_{\varphi_1,\varphi_2}})$. □

We end this section with an observation concerning **R**-Fuchsian embeddings, that is, embeddings which fix an **R**-circle (see also [2]).

Definition 4.9. The **R**-standard elliptic φ_1-configuration is the one where the infinite **R**-circle intersects the y-axis perpendicularly at $y = \sin(\pi/2 - \varphi_1)$.

In this case, we have $\mathbf{R}_1 = \mathbf{R}_{r,\theta,\alpha}$ and $r = \sin(\pi/2 - \varphi_1)$, $\theta = \pi/2$, $\alpha = 0$.

Proposition 4.10. *The* **R***-standard elliptic* φ_1*-configuration is in the same conjugacy class as the* **C***-standard elliptic* $(\varphi_1, -\varphi_1)$*-configuration.*

Proof. The only elliptic elements of **PU**(2,1) which preserve an **R**-circle are elements of that form. $\square$

5. Discrete embeddings

Poincaré's polyhedron theorem is the main tool we use to prove discreteness. A general version, without parabolics, was proved in [1] and we state a version containing only parabolics in [2]. Here we state an appropriate version with both elliptic and parabolic elements.

Let $\{R_i\}$ be a finite collection of finite **R**-circles and $\{S_i\}$ be a collection of **C**-spheres around each of them. Suppose that, pairwise, the **R**-circles *either* intersect at most at one point where the corresponding **C**-spheres intersect tangentially *or* they are linked or intersect twice and the corresponding **C**-spheres intersect in one of the invariant **C**-circles. The intersecting **C**-circles will be called *edges* and the piece of the **C**-surface between two consecutive edges is called a *face* (see Figs. 4 and 5).

Theorem 5.1 (Poincaré polyhedron). *Let* $\{R_i\}$ *be a finite collection of finite* **R***-circles and* $\{S_i\}$ *be a collection of* **C***-spheres around each of them with the hypothesis as above. Suppose that*

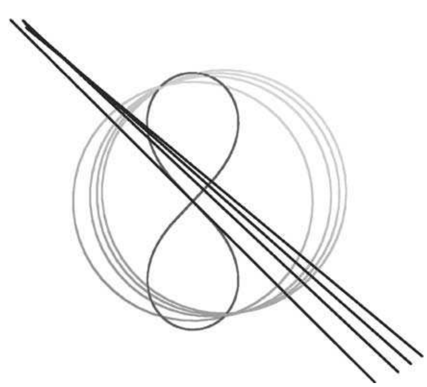


Fig. 4. Four configurations in the family $I_{\pi/3,2\pi/3}(\theta)$: top view showing invariant **C**-circles.

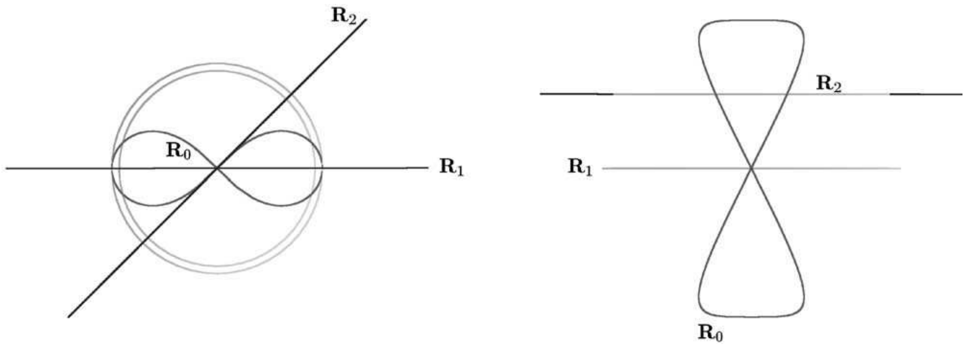


Fig. 5. Standard configuration $\mathbf{C}(\pi/2, 0, \pi/3, 2\pi/3)$: lateral view (right) and top view (left) with invariant $\mathbf{C}$-circles.

(i) *at each edge the angles are of type* $(\pi/q, n\pi/q)$, *where the rotation* π/q *fixes pointwise the edge*,
(ii) *the closure of the unbounded component of the complement of each* $\mathbf{C}$*-sphere containing a face contains all other faces.*

Then the group generated by inversions on each $\mathbf{R}$*-circle is discrete and a fundamental domain is the unbounded component of the complement of the union of all faces.*

The fact that the $\mathbf{C}$-surfaces are unions of $\mathbf{C}$-circles implies that one can extend those surfaces canonically as hypersurfaces in the complex hyperbolic space where they define a "polyhedron". See [1,2] for more details.

5.1. Configurations of the standard and two infinite $\mathbf{R}$*-circles*

Let Γ be the triangular group of type (p, q, ∞), that is, the abstract group presented by

$$\Gamma = \langle \iota_0,\iota_1,\iota_2\colon \iota_0^2 = 1, \iota_1^2 = 1, \iota_2^2 = 1, (\iota_0 \circ \iota_1)^p = 1, (\iota_0 \circ \iota_2)^q = 1\rangle.$$

We want to determine the subspace of $\mathrm{Hom}(\Gamma, \widehat{\mathbf{PU}(2,1)})$ where the image of ι_i are inversions and such that the embedding is injective and discrete. If the image of ι_i are inversions, three $\mathbf{R}$-circles $\mathbf{R}_1,\mathbf{R}_2,\mathbf{R}_0$ are defined. The relations in the presentation imply that $\mathbf{R}_1,\mathbf{R}_2$ can be considered as non-intersecting infinite $\mathbf{R}$-circles and $\mathbf{R}_0$ a finite one. Moreover, in order to prove discreteness, $\mathbf{R}_1,\mathbf{R}_0$ should be equivalent to the $\mathbf{C}$-standard $(\pi/p, n\pi/p)$-configuration and $\mathbf{R}_2,\mathbf{R}_0$ should be equivalent to the $\mathbf{C}$-standard $(\pi/q, m\pi/q)$-configuration, where n, m are integers.

5.2. Standard embeddings of the $(2,3,\infty)$*-triangle group*

In this section, we give examples of standard embeddings in the case of the triangular group of type $(2,3,\infty)$. There are six $\mathbf{C}$-Fuchsian embeddings and one $\mathbf{R}$-Fuchsian embedding.

Observe that in the $\mathbf{C}$-Fuchsian case, $\mathbf{R}_2$ could be embedded as in the $(\pi/3,0)$-configuration, $(\pi/3,\pi/3)$-configuration or $(\pi/3,2\pi/3)$-configuration. $\mathbf{R}_1$ could be as in the $(\pi/2,0)$-configuration or $(\pi/2,\pi/2)$-configuration.

The combination of those cases gives us the six **C**-Fuchsian embeddings. The only **R**-Fuchsian embedding is given by the **R**-standard $\pi/3$-configuration together with the **R**-standard $\pi/2$-configuration.

All those embeddings are discrete and injective. The embeddings will be denoted by their angles. For instance $\mathbf{C}(\varphi_1,\varphi_2,\varphi'_1,\varphi'_2)$ is the **C**-standard embedding with $\mathbf{R}_1$ as in (φ_1,φ_2)-configuration and $\mathbf{R}_2$ as in (φ'_1,φ'_2)-configuration.

5.3. Teichmüller space

Consider the Teichmüller space of the triangle group of type (p,q,∞) in $\widehat{\mathbf{PU}(2,1)}$.

Theorem 5.2.

- *Each embedding* $\mathbf{C}(\pi/p,0,\pi/q,0)$, $\mathbf{C}(\pi/p,\pi/p,\pi/q,\pi/q)$ *or* $\mathbf{C}(\pi/p,0,\pi/q,\pi/q)$ *is isolated in the Teichmüller space.*
- *Each embedding* $\mathbf{C}(\pi/p,0,\pi/q,n\pi/q)$ *or* $\mathbf{C}(\pi/p,\pi/p,\pi/q,n\pi/q)$ *with* $n>1$, *is contained in a real one-dimensional open set of the Teichmüller space.*
- *Each embedding* $\mathbf{C}(\pi/p,m\pi/p,\pi/q,n\pi/q)$ $(n,m>1)$ *is contained in a real two-dimensional open set of the Teichmüller space.*

Proof. We give the idea of the proof, referring to some technical lemmas in the following section.

- A configuration $\mathbf{C}(\pi/p,0,\pi/q,0)$ is isolated in Teichmüller space by applying Theorem 4.4 twice.
- A configuration $\mathbf{C}(\pi/p,\pi/p,\pi/q,n\pi/q)$ can be deformed (if $n>1$) only because of the one-parameter deformation of $\mathbf{R}_2$, $\mathbf{R}_1$ being rigid by Theorem 4.7.
- By Theorem 4.8, we can describe the one parameter family of deformations of $\mathbf{C}(\pi/p,0,\pi/q,n\pi/q)$ $(n>1)$ and the two-parameter family of deformations of $\mathbf{C}(\pi/p,m\pi/p,\pi/q,n\pi/q)$ $(m,n>1)$. In the first case, Theorem 4.4 shows the rigidity of $\mathbf{R}_1$. In the last case we have

$$\alpha = 3\theta - \pi/2 + 2\varphi_1 - \varphi_2$$

and

$$r^2 = -\frac{\cos(2\theta+\varphi_1-\varphi_2)\cos(2\theta+\varphi_1)}{\cos(2\theta+2\varphi_1-\varphi_2)},$$

where $(\varphi_1,\varphi_2)=(\pi/p,m\pi/p)$ or $(\varphi_1,\varphi_2)=(\pi/q,n\pi/q)$ give the coordinates of the infinite **R**-circle $\mathbf{R}_1$ and $\mathbf{R}_2$ as $\mathbf{R}_{r(\theta),\theta,\alpha(\theta)}$. Of course, if $p=2$ we use those formulas only for $\mathbf{R}_2$, $\mathbf{R}_1$ being rigid (see Fig. 6).

Using Poincaré's theorem we should find three **C**-spheres S_0,S_1,S_2 invariant, respectively, by $I_0,I_1(\theta_1),I_2(\theta_2)$, the inversions in the three **R**-circles such that the intersection of S_1 and S_2 is the point at infinity, the intersection of S_1 and S_0 is one of the invariant **C**-circles by I_0 and $I_1(\theta_1)$ and the intersection of S_2 and S_0 is one of the invariant **C**-circles by I_0 and $I_2(\theta_2)$. The surfaces will depend continuously on θ_1 and θ_2, but we will not write explicitly the angles in order to simplify notations.

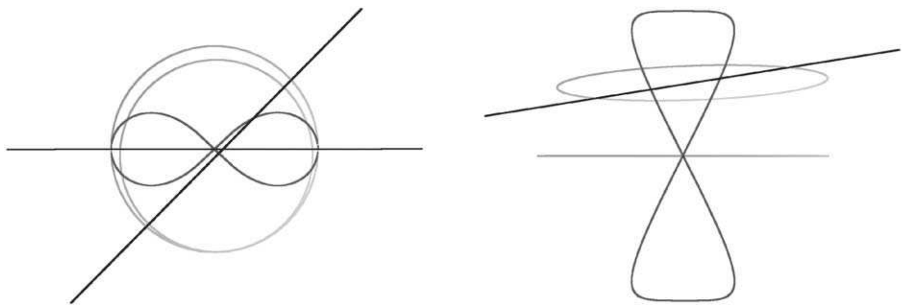


Fig. 6. Deformed configuration; lateral view (right) and top view (left) with invariant **C**-circles.

The **C**-circle $\mathbf{C}_1$, invariant by I_0 and $I_1(\theta_1)$ is the one with center determined by $r(\theta_1)$ in the surface of centers. Call that center M_1 and M_2 the center of the **C**-circle $\mathbf{C}_2$.

For each one of the three surfaces we need to define the **C**-circles composing them and that will be done by describing a curve of centers of **C**-circles with an appropriate function of those centers describing their *radii*.

5.3.1. Standard fundamental domain

The invariant surface for $\mathbf{R}_0$ will be given by its axis $\rho(\theta)$ with $\theta_1 \leqslant \theta \leqslant \theta_2$ (see Lemma 6.3).

In the standard embedding, M_1 is on the vertical axis with height $\cos(\pi/p)$ and the **C**-circle $\mathbf{C}_1$ is a horizontal **C**-circle with center in M_1 of *radius* $\sqrt{\sin(\pi/p)}$. A fundamental domain in that case is obtained by defining S_1 to be the union of concentric **C**-circles, analogous for S_2 and S_0 to be the union of **C**-circles with centers on the vertical axis from M_2 to M_1.

5.3.2. Deformation of the standard fundamental domain

Let $N_1 \in \mathbf{R}_1$ and $N_2 \in \mathbf{R}_2$ whose projections are the intersection of projections of $\mathbf{R}_1$ and $\mathbf{R}_2$ on the z-plane. For $\mathbf{R}_1$, we will take centers in the segment $[M_1, N_1]$ with appropriate *radii*. Then we will complete the surface by a union of **C**-circles of centers N_1. We proceed analogously with $\mathbf{R}_2$ (see Lemma 6.4) (see Fig. 7).

1. *Chimney.* If $\mathbf{C}_1$ and $\mathbf{C}_2$ will be near the corresponding **C**-circle of the standard embedding, we will be able to choose an axis such that the corresponding **C**-circles will be above the plane determined by $\mathbf{C}_1(P(M_1))$ and below the plane determined by $\mathbf{C}_2(P(M_2))$.

2. *Beginning of* S_1 *and* S_2. We then may choose concentric circles centered on M_1 and M_2 with increasing *radii* until their projections are large containing the projection of S_0. This assures that the families S_1 and S_2 do not intersect the family S_0. If the deformation is small enough we are, moreover, certain that the families S_1 and S_2 do not intersect (see Proposition 6.5).

3. *Middle of* S_1 *and* S_2. We should then move the centers of the **C**-circles from M_1 to N_1 on $\mathbf{R}_1$ and from M_2 to N_2 on $\mathbf{R}_2$. Again, if the deformation is small enough this can be done in such a way that there will be no intersection between the families S_1 and S_2 (see Lemma 6.4).

4. *End of* S_1 *and* S_2. Finally, from the points N_1 and N_2 we complete the construction of S_1 and S_2 with concentric **C**-circles. They are parallel.

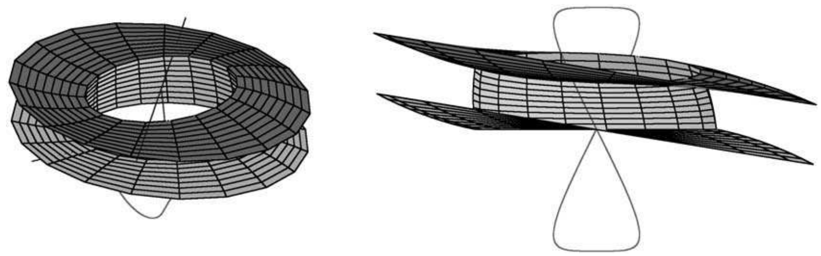


Fig. 7. Two views of the fundamental domain.

The final construction is explicit in Proposition 6.5. Observe that, for the standard case, parts 2 and 3 are empty. □

Remark 5.3. Figs. 6 and 7 show the deformation of the triangle group corresponding to the configuration $\mathbf{C}(\pi/2,0,\pi/3,2\pi/3)$. That group contains the modular group as an index two subgroup.

Remark 5.4. Analogously, combining present methods and those of [1], a triangle group of type (p,∞,∞) has Teichmüller neighborhoods of dimension two and three if $p>2$. The triangle group of type $(2,\infty,\infty)$ has Teichmüller neighborhoods of dimension two.

Remark 5.5. We did not determine explicitly the complete Teichmüller space. We conjecture that it contains the points of the configuration space where the **C**-circles $\mathbf{C}_1$ and $\mathbf{C}_2$ are not linked. In particular, using the methods of this paper one can prove that the **R**-Fuchsian standard embeddings $\mathbf{R}(\pi/2,\pi/3)$ are in the same component as $\mathbf{C}(\pi/3,-\pi/3,\pi/2,-\pi/2)$.

6. Technical Lemmas

Lemma 6.1. *Let $P(M_1)$ be the contact plan at $M_1=(r_1,\theta_1,t_1)$. Let C_2 be a* **C**-*circle of center $M_2=(r_2,\theta_2,t_2)$ and radius R_2. For any point M in C_2 (see notations in Proposition* 3.6)

$$|h+4S|-2dR_2 \leqslant \sqrt{1+4r_1^2}\, d(M,P(M_1)) \leqslant |h+4S|+2dR_2.$$

Proof. For $M=(x,y,z)$, we have $d(M,P(M_1))\sqrt{1+4r_1^2}=|z-t_1+2r_1(\cos\theta_1 y-\sin\theta_1 x)|$. Here $x=r_2\cos\theta_2+R\cos\varphi$, $y=r_2\sin\theta_2+R\sin\varphi$, $z=t_2+2r_2R\sin(\varphi-\theta_2)$ so

$$d(M,P(M_1))=\frac{1}{\sqrt{1+4r_1^2}}\,|t_2-t_1+2r_1r_2\sin(\theta_2-\theta_1)+2R(r_2\sin(\varphi-\theta_2)-r_1\sin(\varphi-\theta_1))|.$$

But $|r_2\sin(\varphi-\theta_2)-r_1\sin(\varphi-\theta_1)|\leqslant d$ so the announced result. □

Lemma 6.2 (Upper chimney). *Let $M_0=(0,\theta_0,\sin 2\theta_0)$ and $M_2=(r_2,\theta_2,\sin 2\theta_2)$ with $-\pi/4<\theta_1<\theta_2<\pi/4$ and let $r_2=a\sin(\theta_2-\theta_0)$. Then the axis $r(\theta)=a\sin(\theta-\theta_0)$, $\theta_0\leqslant\theta\leqslant\theta_2$ is a good* *axis when $r_2^2 < m_2(\theta_0,\theta_2)$. Furthermore, for $\theta_0 \leqslant \theta \neq \theta' \leqslant \theta_2$, the* **C***-circle $C(\theta)$ does not intersect the plane containing the* **C***-circle $C(\theta')$.*

Proof. Let $M = (r(\theta), \theta, \sin 2\theta)$, $M' = (r(\theta'), \theta', \sin 2\theta')$ with $r(\theta) = a\sin(\theta - \theta_0)$. We have

$$\begin{aligned}4d^2R(\theta)^2 - (h+4S)^2 &= 4a^2\sin(\theta-\theta')^2(a^2\sin(\theta-\theta_0)^2+\cos 2\theta)\\ &\quad - (2\sin(\theta'-\theta)(\cos(\theta'+\theta)+a^2\sin(\theta-\theta_0)\sin(\theta'-\theta_0)))^2\\ &= 4\sin(\theta-\theta')^2(a^2(a^2\sin(\theta-\theta_0)^2+\cos(2\theta)) - (\cos(\theta+\theta')\\ &\quad + a^2\sin(\theta-\theta_0)\sin(\theta'-\theta_0))^2).\end{aligned}$$

If $a^2 < a^2\sin(\theta'-\theta_0)^2 + \cos(2\theta')$ then $a^2\cos(2\theta_0) < 1$ and

$$\begin{aligned}&a^2(a^2\sin(\theta-\theta_0)^2+\cos 2\theta) - (\cos(\theta+\theta')+a^2\sin(\theta-\theta_0)\sin(\theta'-\theta_0))^2\\ &\quad\leqslant (a^2\sin(\theta'-\theta_0)^2+\cos 2\theta')(a^2\sin(\theta-\theta_0)^2+\cos 2\theta) - (\cos(\theta+\theta')\\ &\qquad + a^2\sin(\theta-\theta_0)\sin(\theta'-\theta_0))^2\\ &\quad= (a^2\cos 2\theta_0 - 1)\sin(\theta-\theta')^2 < 0.\end{aligned}$$

For instance, we must have $a^2 < \min(\cos 2\theta_0, \cos 2\theta_2/\cos(\theta_2-\theta_0)^2) = m_2(\theta_0,\theta_2)/\sin(\theta_2-\theta_0)^2$. □

We thus deduce

Lemma 6.3 (Chimney). *Let $M_1 = (r_1, \theta_1, \sin 2\theta_1)$ and $M_2 = (r_2, \theta_2, \sin 2\theta_2)$ with $-\pi/4 < \theta_1 < \theta_2 < \pi/4$. If r_1 and r_2 are small enough then there exists a good axis from M_1 to M_2. Furthermore, corresponding* **C***-circles are between the contact planes $P(M_1)$ at M_1 and $P(M_2)$ at M_2.*

Proof. As in Lemma 6.2, we build a lower chimney from $M_1 = (r_1, \theta_1, \sin 2\theta_1)$ and $M_0 = (0, \theta_0, \sin 2\theta_0)$ with $-\pi/4 < \theta_1 < \theta_0 < \pi/4$ under the condition:

$$r_1^2 \leqslant m_1(\theta_0, \theta_1) = \min(\cos 2\theta_0, \cos 2\theta_1/\cos(\theta_1-\theta_0)^2)\sin(\theta_1-\theta_0)^2.$$

Let us consider now $\theta_0 = \frac{1}{2}(\theta_1 + \theta_2)$. If $r_1^2 < m_1(\theta_0, \theta_1)$ and $r_2^2 < m_2(\theta_0, \theta_2)$ the concatenation of the lower and upper chimney gives us the announced result. □

Observe that in this case for any **C**-circle, we have $R^2 = r^2 + \cos 2\theta \leqslant \max(r_1^2, r_2^2) + 1$.

Lemma 6.4. *Let $\mathbf{R}_1 = \mathbf{R}_{r_1,\theta_1,\alpha_1}$ and $\mathbf{R}_2 = \mathbf{R}_{r_2,\theta_2,\alpha_2}$ be two infinite* **R***-circles and let C_1, C_2 be two* **C***-circles centered at M_1 and M_2 with radii R_1 and R_2. Let d be the horizontal distance between M_1 and M_2. If*

$$\frac{2d}{|\sin(\alpha_1-\alpha_2)|}\left(|r_1| + |r_2| + 2\max(R_1, R_2) + \frac{2d}{|\sin(\alpha_1-\alpha_2)|}\right) \leqslant |\sin(2\theta_2 - 2\theta_1)|$$

then there exist an invariant surface S_1 of $\mathbf{R}_1$ and an invariant surface S_2 for $\mathbf{R}_2$ that do not intersect.

Proof. Let $M=(r,\theta,0)$ the intersection point of projections of $\mathbf{R}_1$ and $\mathbf{R}_2$. It is also the projection of N_1 and N_2. We have

$$r\sin(\theta-\alpha_1)=r_1\sin(\theta_1-\alpha_1),\quad r\sin(\theta-\alpha_2)=r_2\sin(\theta_2-\alpha_2).$$

$$d(N_1,M_1)=\left|\frac{r_2\sin(\theta_2-\alpha_2)-r_1\sin(\theta_1-\alpha_2)}{\sin(\alpha_1-\alpha_2)}\right|\leqslant\frac{d}{|\sin(\alpha_1-\alpha_2)|},$$

$$d(N_2,M_2)=\left|\frac{r_2\sin(\theta_2-\alpha_1)-r_1\sin(\theta_1-\alpha_1)}{\sin(\alpha_1-\alpha_2)}\right|\leqslant\frac{d}{|\sin(\alpha_1-\alpha_2)|}.$$

Consider the family of $\mathbf{C}$-circles centered at $Q_1=(r,\theta,Z(\theta))\in[M_1,N_1]\subset\mathbf{R}_1$, of *radius* $R_1+\lambda_1 d(Q_1,M_1)\leqslant R$. If $\lambda_1>1$ then this family is not linked because projections of $\mathbf{C}$-circles are not. For any point Q_1 of this family of $\mathbf{C}$-circles we have

$$d(Q_1,P(M_1))\leqslant 2d(N_1,M_1)(R_1+\lambda_1 d(Q_1,M_1))\leqslant 2\frac{dR}{|\sin(\alpha_1-\alpha_2)|}.$$

Consider the family of $\mathbf{C}$-circles centered at $Q_2=(r,\theta,Z(\theta))\in[M_2,N_2]\subset\mathbf{R}_2$. of *radius* $R_2+\lambda_2 d(Q_2,M_2)\leqslant R$ with $\lambda_2>1$. For any point Q_2 of this family. We have (in $d(Q_2,P(M_1))$)

$$|h|-|t_2-t_1|\leqslant 2|r_2|\,|\sin(\alpha_2-\theta_2)|d(N_2,M_2),$$

$$|4S|\leqslant 2|r_1|\,|\sin(\alpha_1-\theta_1)|d(Q_2,M_1)$$

so

$$d(Q_2,P(M_1))\geqslant|t_2-t_1|-2\frac{(|r_1|+|r_2|+R)d}{|\sin(\alpha_1-\alpha_2)|}.$$

We choose λ_1 and λ_2 such that $\lambda_i>1$ and

$$R=R_1+\lambda_1 d(N_1,M_1)=R_2+\lambda_2 d(N_2,M_2)<\max(R_1,R_2)+d/|\sin(\alpha_1-\alpha_2)|.$$

We then have $d(Q_1,P(M_1))<d(Q_2,P(M_2))$ for any points $(Q_1,Q_2)\in S_1\times S_2$. $\square$

Proposition 6.5 (Non-intersecting invariant surfaces). *Let* $\mathbf{R}_1=\mathbf{R}_{r_1,\theta_1,\alpha_1}$ *and* $\mathbf{R}_2=\mathbf{R}_{r_2,\theta_2,\alpha_2}$ *be two infinite* $\mathbf{R}$*-circles with* $\sin(\alpha_1-\alpha_2)\neq 0$. *If* r_1 and r_2 *are small enough then there exist three invariant surfaces* S_0,S_1,S_2 *for corresponding inversions* I_0,I_1,I_2.

Proof. Let $\theta_0=\frac{1}{2}(\theta_1+\theta_2)$. Suppose we have $r_1^2\leqslant m_1(\theta_0,\theta_1)$ and $r_2^2\leqslant m_2(\theta_0,\theta_2)$.666666

(1) We first build an invariant surface S_0, using the chimney lemma. Its axis is given by the concatenation of the two axis:

$$r(\theta)=r_1\frac{\sin(\theta_0-\theta)}{\sin(\theta_0-\theta_1)},\quad\theta_1\leqslant\theta\leqslant\theta_0,\qquad r(\theta)=r_2\frac{\sin(\theta-\theta_0)}{\sin(\theta_2-\theta_0)},\quad\theta_0\leqslant\theta\leqslant\theta_2,$$

Radii of corresponding $\mathbf{C}$-circle are bounded by $\sqrt{1+r_1^2}$ and $\sqrt{1+r_2^2}$, respectively. S_0 is inside the cylinder $d(M,M_1)=1+2r_1=R_1$ and the cylinder $d(M,M_2)=1+2r_2=R_2$.

(2) We then build a family of concentric **C**-circles at M_1 with *radius* from $\sqrt{r_1^2+\cos 2\theta_1}$ to R_1. This family of **C**-circle is part of S_1 and is a subset of $P(M_1)$ so does not intersect S_0. In the same way, we build a family of concentric **C**-circles with *radius* from $\sqrt{r_2^2+\cos 2\theta_2}$ to R_2.
If $|r_1 r_2 \sin(\theta_1-\theta_2)| + d(1+2\max(|r_1|,|r_2|)) < \frac{1}{2}|\sin 2\theta_1 - \sin 2\theta_2|$, these family do not intersect.

(3) Using Lemma 6.4, we then build two family of invariant **C**-circles for I_1 and I_2 that do not intersect if

$$\frac{2d}{|\sin(\alpha_1-\alpha_2)|}\left(|r_1|+|r_2|+2+4\max(|r_1|,|r_2|)+\frac{2d}{|\sin(\alpha_1-\alpha_2)|}\right) < |\sin(2\theta_1-2\theta_2)|.$$

All the **C**-circles we have already built are inside the cylinder $d(M_0,M)=R$.

(4) We then consider the two concentric families of **C**-circle centered at N_1 and N_2 with growing *radius* from R to infinity. They are lying in the two parallel contact planes at N_1 and N_2. □

References

[1] E. Falbel, P.-V. Koseleff, Flexibility of the ideal triangle group in complex hyperbolic geometry, Topology 39 (2000) 1209–1223.
[2] E. Falbel, V. Zocca, A Poincaré's polyhedron theorem for complex hyperbolic geometry, J. Reine Angew. Math. 516 (1999) 133–158.
[3] W. Goldman, Representations of Fundamental Groups of Surfaces, Geometry and Topology, Proceedings, University of Maryland 1983–1984, Lecture Notes in Mathematics, Vol. 1167, 1985, pp. 95–117.
[4] W. Goldman, Complex Hyperbolic Geometry, Oxford Mathematical Monographs, Oxford Science Publications, Oxford, 1999.
[5] W. Goldman, J.J. Millson, Local rigidity of discrete groups acting on complex hyperbolic space, Invent. Math. 88 (1987) 495–520.
[6] W. Goldman, J. Parker, Complex hyperbolic ideal triangle groups, J. Reine Angew. Math. 425 (1992) 71–86.
[7] N. Gusevskii, J. Parker, Representations of free Fuchsian groups in complex hyperbolic space, Topology 39 (2000) 33–60.
[8] N.J. Hitchin, Lie groups and Teichmüller space, Topology 31 (3) (1992) 449–473.
[9] A. Korányi, H.M. Reimann, Quasiconformal mappings on the Heisenberg group, Invent. Math. 80 (1985) 309–338.
[10] J. Parker, Private communication.
[11] R. Schwartz, Ideal triangle groups, denoted tori and numerical analysis, preprint, 1997.
[12] D. Toledo, Representations of surface groups in complex hyperbolic space, J. Differential Geom. 29 (1989) 125–133.

A circle of Modular Groups in $\mathbf{PU}(2,1)$

E. FALBEL, P.-V KOSELEFF
Mathematical Research Letters 9 (2002), pp. 379-394

Mathematical Research Letters **9**, 1–14 (2002)

A CIRCLE OF MODULAR GROUPS IN PU(2,1)

E. Falbel and P.-V. Koseleff

Abstract. We prove that there exists a circle of discrete and faithful embeddings of the triangle group of type $(2, 3, \infty)$ in the automorphisms group of complex hyperbolic space. The proof is obtained by a construction of a fundamental domain using **C**-spheres.

1. Introduction

Let Γ be the triangle group of type (p, q, ∞), that is, the abstract group presented by

$$< \iota_0, \iota_1, \iota_2 : \iota_0^2 = 1, \iota_1^2 = 1, \iota_2^2 = 1, (\iota_0 \circ \iota_1)^p = 1, (\iota_0 \circ \iota_2)^q = 1 > .$$

By an *embedding* of Γ in $\widehat{\mathbf{PU}(2,1)}$ (containing the holomorphic and the anti-holomorphic transformations), the isometry group of the two dimensional complex ball (with the natural metric invariant under biholomorphisms), we will consider an homomorphism such that ι_i are mapped to anti-holomorphic generators and such that $\iota_1 \circ \iota_2$ is parabolic. In this paper we prove:

Theorem 1.1. *There exists a circle of discrete faithful embeddings of the triangle group* $(2, 3, \infty)$ *in* $\widehat{\mathbf{PU}(2,1)}$. *Up to conjugation in* $\widehat{\mathbf{PU}(2,1)}$ *the family is reduced to a quotient by the dihedral group* $\mathbf{Z_2} \times \mathbf{Z_2}$. *Moreover the family contains embeddings fixing a complex geodesic and embeddings fixing a totally real totally geodesic plane.*

Embeddings of triangle groups of type (p, q, ∞) in the neighborhood of an embedding fixing a complex geodesic were analyzed previously in [FK2]. A family connecting embeddings preserving a real and a complex geodesic for the triangle modular group was also obtained in [GuP1] independently, but there, fundamental domains were constructed for a subgroup of the triangle group. We prove discreteness of the embeddings by constructing explicitly fundamental domains using **C**-spheres (see [FZ, FK1]).

Received August 24, 2001.

The technique used in this paper is also sufficient to prove the same result for triangle groups (p,q,∞) for $p \leq q \leq 4$. In the case $3 \leq p,q$ we obtain an open neighborhood of a circle in the set of embeddings in $\widehat{\mathbf{PU}(2,1)}$. On the other hand, for other triangle groups the technique proves the existence of deformations of embeddings fixing either a complex geodesic or a totally real one.

2. Complex hyperbolic space and its boundary

In this section and the following we collect general results about complex hyperbolic space. As references, we use [G2], [FK1] and [FK2].

In dimension one, the disc and the half-plane are related by a Cayley transform. In dimension two, complex hyperbolic space

$$H^2_{\mathbf{C}} = \{\ w \in \mathbf{C}^2 \ :\ \ |w_1|^2 + |w_2|^2 < 1\ \ \}$$

is biholomorphic to

$$V = \{\ z \in \mathbf{C}^2 \ :\ \ \mathrm{Im}\,(z_2) > |z_1|^2\ \ \},$$

using the Cayley transform

$$C : (w_1, w_2) \mapsto \left(\frac{iw_1}{1+w_2}, i\frac{1-w_2}{1+w_2}\right).$$

The *Heisenberg group* $\mathbf{H}$ is the set of pairs $(z,t) \in \mathbf{C} \times \mathbf{R}$ with the product

$$(z,t)\cdot(z',t') = (z+z', t+t'+2\mathrm{Im}\, z\overline{z}').$$

We identify the boundary of V with the Heisenberg group via the map

$$(z_1, z_2) \mapsto (z_1, \mathrm{Re}\,(z_2)).$$

The distribution obtained by translating the $t=0$ plane at the origin is of contact type and makes the Heisenberg group a homogeneous contact manifold. Moreover, a homogeneous conformal structure on the distribution can be defined by translating the conformal class defined by the flat metric on the $t=0$ plane at the origin.

It is well known that the boundary S^3 of the complex hyperbolic space can be identified to the one point compactification $\overline{\mathbf{H}}$ of the Heisenberg group. The group of all conformal transformations of $\overline{\mathbf{H}}$ is isomorphic to $\widehat{\mathbf{PU}(2,1)}$ (see [G2]).

There are two kinds of totally geodesic submanifolds of real dimension 2 in $H^2_{\mathbf{C}}$: *complex geodesics* (represented by $H^1_{\mathbf{C}} \subset H^2_{\mathbf{C}}$) and *totally real geodesic 2-planes* (represented by $H^2_{\mathbf{R}} \subset H^2_{\mathbf{C}}$). Each of these totally geodesic submanifolds is a model of the hyperbolic plane.

We define $\mathbf{C}$-*circles* to be the intersections of S^3 with the boundaries of totally geodesic complex submanifolds $H^1_{\mathbf{C}}$ in $H^2_{\mathbf{C}}$. Analogously, we define $\mathbf{R}$-*circles* to be the intersections of S^3 with the boundaries of totally geodesic totally real submanifolds $H^2_{\mathbf{R}}$ in $H^2_{\mathbf{C}}$.

Definition 2.1. *The inversion on an* $\mathbf{R}$*-circle is the non-trivial element of* $\widehat{\mathbf{PU}(2,1)}$ *that fixes it point-wise.*

There is, then, a one-to-one correspondence between inversions and $\mathbf{R}$-circles. For instance the transformation $\hat{m}(z,t) = (\overline{z}, -t)$ on the Heisenberg group is the inversion that fixes point-wise the $\mathbf{R}$-circle $\mathrm{Im}\,(z) = 0$.

We will also consider points to be $\mathbf{C}$-circles and refer to them as degenerate $\mathbf{C}$-circles. Any two points (possibly coincident) in the Heisenberg group determine a unique $\mathbf{C}$-circle (possibly degenerate) containing them (see [G2]). One can easily show that the $\mathbf{C}$-circles that are invariant under a given inversion are precisely those that intersect twice the associated $\mathbf{R}$-circle. As a permutation of two points gives the same $\mathbf{C}$-circle, the set of invariant $\mathbf{C}$-circles is the Möbius band $S^1 \times S^1/\mathbf{Z}_2$.

Given an $\mathbf{R}$-circle, a union of invariant $\mathbf{C}$-circles that is homeomorphic to a sphere is called a $\mathbf{C}$-*sphere* (see [FZ]). Fundamental domains in this paper will be bounded by (pieces of) $\mathbf{C}$-spheres.

2.1. R-circles and C-circles in the Heisenberg model. We denote by I_0 the inversion on $\mathbf{R}_0$ (see [G2] and [FK1] for details)

$$\mathbf{R}_0 \; : \; r^2 + iz = -e^{-2i\theta}.$$

$\mathbf{R}_0$ is given in cylindrical coordinates by $r = \sqrt{-\cos(2\theta)}, z = \sin(2\theta)$. The following lemma describes in cylindrical coordinates the homogeneous contact distribution in the Heisenberg group.

Lemma 2.2. *The contact plane at* $M = (r, \theta, z)$ *is :* $z - Z - 2r\,(\cos(\theta)\,Y - \sin(\theta)\,X) = 0$*.*

Any infinite $\mathbf{R}$-circle lies in the contact plane at each of its points.

In the Heisenberg model, $\mathbf{C}$-circles are either vertical lines or else ellipses that belong to the contact plane of their centers and whose projections onto the z-plane are circles.

The Möbius band of all invariant $\mathbf{C}$-circles described in the previous section can be parametrized by the surface defined by all their centers. In order to obtain that surface we need to impose that the $\mathbf{C}$-circle having center (X, Y, Z)

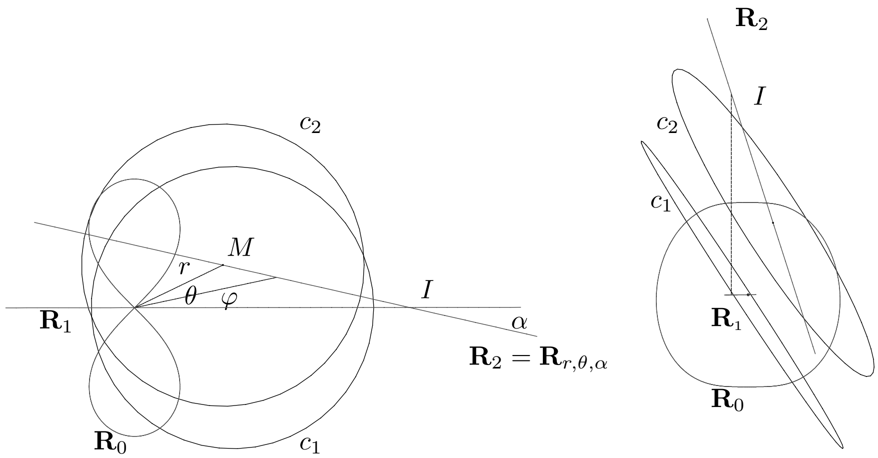
and radius R intersects twice the $\mathbf{R}$-circle $\mathbf{R}_0$. After some calculations, one obtains:

Lemma 2.3 (see [FK1]). *Invariant* $\mathbf{C}$*-circles under the inversion on* $\mathbf{R}_0$ *are centered in the surface of centers* $\{z = \sin(2\theta)\}$ *and have radius* $R^2 = r^2 + \cos(2\theta)$. *If* $r^2 + \cos(2\theta) \leq 0$ *we say that the* $\mathbf{C}$*-circle is imaginary.*

As a $\mathbf{C}$-sphere around $\mathbf{R}_0$ is union of invariant $\mathbf{C}$-circles, one can describe it by giving a path (called an axis) in the surface of centers. But the condition that it is homeomorphic to a sphere imposes certain restrictions to that path. One of the simplest sufficient conditions is to impose that the $\mathbf{C}$-circles in the family defined by the axis be non-intersecting and non-linked pairwise (see section 3.1).

Definition 2.4. $\mathbf{R}_{r,\theta,\alpha}$ *is the infinite* $\mathbf{R}$*-circle passing through* $M = (r, \theta, \sin(2\theta))$ *whose projection on the* z*-plane is the line of slope* $\tan(\alpha)$.

Observe that $\mathbf{R}_{r,\theta,\alpha}$ is horizontal if and only if $\alpha = \theta \mod \pi$.



Left: Projection on the z-plane of the configuration of $\mathbf{R}_0$, $\mathbf{R}_1$ and $\mathbf{R}_2(\theta)$ for the modular group showing the invariant $\mathbf{C}$-circles c_1 and c_2. In the picture we fixed $\theta = \pi/7$. $\mathbf{R}_2(\theta)$ has polar coordinates ρ and φ.

Right: A different projection of the configuration.

2.2. Configurations. Let ι_0, ι_1 and ι_2 be the standard generators of the triangle group $(2, 3, \infty)$. Consider an embedding $r : (2, 3, \infty) \to \widehat{\mathbf{PU}(2,1)}$ such that $r(\iota_i) = I_i$ are inversions fixing three $\mathbf{R}$-circles $\mathbf{R}_i$ and such that $I_1 \circ I_2$ is parabolic.

Observe that $I_1 \circ I_0$ is an elliptic element of order 2 and $I_2 \circ I_0$ is an elliptic element of order 3. $I_1 \circ I_0$ is conjugate to an element in $U(2)$. Its eigenvalues may be ± 1. Similarly $I_2 \circ I_0$ could have eigenvalues $1, e^{2\pi/3}, e^{-2\pi/3}$. There are 6 possible families of embeddings.

As $\mathbf{R}$-circles are tangent to the contact distribution, one can define an angle θ between them whenever they intersect. In that case, the eigenvalues of the composition of inversions will be 1 and 2θ. If they do not intersect but are linked one can turn one of them by an angle φ around an invariant $\mathbf{C}$-circle under both inversions until they intersect defining a new angle θ. In that case, the eigenvalues of the composition will be 2φ and 2θ (see [FZ] for details). In order to parameterize the configurations one needs to impose for each pair of $\mathbf{R}$-circles a fixed pair of angles.

In [FK2], the family of embeddings such that $I_1 \circ I_0$ have both eigenvalues ± 1 and such that $I_2 \circ I_0$ have both eigenvalues $e^{2\pi/3}, e^{-2\pi/3}$ are called of type $(\pi/2, -\pi/2, \pi/3, -\pi/3)$. It contains two distinguished discrete and faithful representations, namely one preserving a complex line and one preserving a real line. A parameterization of those embeddings is given by the following proposition.

Consider in the Heisenberg group $\mathbf{R}_0$ to be the standard $\mathbf{R}$-circle, $\mathbf{R}_1$ to be the infinite $\mathbf{R}$-circle given by the x-axis and $\mathbf{R}_2$ to be the family of infinite $\mathbf{R}$-circles $\mathbf{R}_2(\theta) = \mathbf{R}_{r,\theta,\alpha}$ where

$$\alpha = -\pi/2 + 3\theta, r^2 = -\frac{\cos(2\theta - \pi/3)\cos(2\theta + \pi/3)}{\cos(2\theta)} = -\frac{\cos(6\theta)}{4\cos^2(2\theta)}$$

for $\frac{\pi}{12} \leq \theta \leq \frac{\pi}{6}$. We denote by I_0 the inversion fixing $\mathbf{R}_0$, by I_1 the inversion fixing the $\mathbf{R}_1$ and by $I_2(\theta)$ the inversion on the $\mathbf{R}$-circle $\mathbf{R}_2(\theta)$.

Theorem 2.5 (FK2)**.** *The representations r of the triangle group $(2, 3, \infty)$ with $r(\iota_0) = I_0$, $r(\iota_1) = I_1$ and $r(\iota_2) = I_2(\theta)$ form a connected component of the representation space up to conjugations in $\widehat{\mathbf{PU}(2,1)}$. Moreover if $\theta = \pi/12$ the representation fixes a complex line (a $\mathbf{C}$-fuchsian representation) and if $\theta = \pi/6$, the representation fixes a totally real totally geodesic plane (a $\mathbf{R}$-fuchsian representation).*

A circle of representations in $\widehat{\mathbf{PU}(2,1)}$ is obtained by conjugating the group by the inversions fixing the x-axis and the y-axis. The index two subgroup generated by the order two and order three elements will have, a fortiori, a circle of representations in $\mathbf{PU}(2,1)$. The space of representations modulo conjugation in $\mathbf{PU}(2,1)$ is obtained by the quotient of this circle by a $\mathbf{Z}_2$-action (a rotation by π).

3. Families of non-linked C-circles

3.1. Non-linking condition. The linking condition can be formulated as follows.

Let C_1 and C_2 be two $\mathbf{C}$-circles and M_1, M_2 be their centers with parameters (r_1, θ_1, z_1) and (r_2, θ_2, z_2). Their radius are R_i. They are not linked if and only if

$$E = d^4 - 2(R_1^2 + R_2^2)d^2 + (R_1^2 - R_2^2)^2 + (h + 4S)^2 > 0,$$

where $h = z_2 - z_1$ and $S = 1/2(x_1y_2 - x_2y_1)$, $d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$ (see [FK1]).

3.2. Good axis in the surface of centers. Let $M(\theta)$ be lying on the surface of centers of $\mathbf{R}_0$, that is $M(\theta) = (r, \theta, \sin(2\theta))$. It defines an invariant $\mathbf{C}$-circle $C(\theta)$ for $\mathbf{R}_0$ with radius $R^2(\theta) = r^2(\theta) + \cos(2\theta)$.

Lemma 3.1. *Let $r_0 \leq 1$. The curve $M(\theta) = (r_0, \theta, \sin(2\theta))$ in the surface of centers defines a family of invariant non-linked $\mathbf{C}$-circles for $\mathbf{R}_0$, when $\theta \in [-\pi/2, \pi/2]$.*

Proof. Let $r_1 = r_2 = r_0$, we get $1/4\,E$

$$\begin{aligned}
&= -r_0^2(\cos(2\theta_1) + \cos(2\theta_2) - 2\cos(\theta_1 + \theta_2)) + \sin(\theta_1 - \theta_2)^2 \\
&= (1 - \cos(\theta_1 - \theta_2))\left((\cos(\theta_1 - \theta_2) + r_0^2\cos(\theta_1 + \theta_2)) + (r_0^2\cos(\theta_1 + \theta_2) + 1)\right) \\
&= (1 - \cos(\theta_1 - \theta_2))\left(\cos(\theta_1)\cos(\theta_2)(1 + 2r_0^2) + \sin(\theta_1)\sin(\theta_2)(1 - 2r_0^2) + 1\right).
\end{aligned}$$

But

$$\cos(\theta_1)\cos(\theta_2)(1 + 2r_0^2) + \sin(\theta_1)\sin(\theta_2)(1 - 2r_0^2) + 1 \geq \min(2r_0^2, 2 - 2r_0^2) \geq 0.$$

□

We call such a curve a good axis (cf. [FK1]).

3.3. Dilated C-circles. We construct a two parameter family of unlinked $\mathbf{C}$-circles using dilations $l_\lambda : (z, t) \mapsto (\lambda z, \lambda^2 t)$ in the Heisenberg group.

Lemma 3.2. *Let $r(\theta), \theta \in I$ be a curve in the surface of centers such that corresponding $\mathbf{C}$-circles are not linked (*good axis *in* [FK1]*). Let $C_\lambda(\theta) = l_\lambda(C(\theta))$ be the two-parameter family of dilated $\mathbf{C}$-circles $C(\theta)$ by dilation l_λ. For any*

$\lambda_1, \lambda_2 > 0$ and $(\theta_1, \theta_2) \in I \times I$. If $r(\theta_1)r(\theta_2)\cos(\theta_1+\theta_2) \geq -1$, then $C_{\lambda_1}(\theta_1)$ and $C_{\lambda_2}(\theta_2)$ are not linked.

Remark 3.3. *Observe that the surfaces $C_\lambda(\theta)$ (constant λ) are disjoint. Moreover, by the previous lemma, as $r(\theta) = r_0 \leq 1$ is a good axis, the conclusion of the lemma is true in that case.*

Proof. Let $C(\theta)$ be a $\mathbf{C}$-circle of center $M(\theta)$ and radius $R(\theta)$, then $C_\lambda(\theta)$ has center $M_\lambda(\theta)$ and radius λR. Linking formula for $C_{\lambda_1}(\theta_1)$ and $C_{\lambda_2}(\theta_2)$ is

$$\begin{aligned} F &= \lambda_1^4 + \lambda_2^4 - 2\lambda_1^2\lambda_2^2(2r_2^2\cos(2\theta_1) + 2r_1^2\cos(2\theta_2) + \cos(2\theta_1 - 2\theta_2)) \\ &\qquad + 4\lambda_1\lambda_2 r_1 r_2(\lambda_1^2 + \lambda_2^2)\cos(\theta_1+\theta_2) \\ &= \lambda_1^2\lambda_2^2 E + (\lambda_1 - \lambda_2)^2\left(\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2(1 + 2r_1r_2\cos(\theta_1+\theta_2))\right) \\ &\geq 0 \end{aligned}$$

when $E \geq 0$ and $r_1 r_2 \cos(\theta_1+\theta_2) \geq -1$. □

3.4. Invariant C-circles for an infinite R-circle.

Lemma 3.4. *Let $C(\varphi), \varphi_1 \leq \varphi \leq \varphi_2$ be a family of $\mathbf{C}$-circle of radius $R(\varphi)$ centered at $M(\varphi)$ which projects to a line in the the xy-plane . Suppose that $\varphi \mapsto R(\varphi)$ is differentiable and let $d(\varphi)$ be the horizontal distance between $M(\varphi)$ and $M(\varphi_2)$. If $\left[\frac{\mathrm{d}R^2(\varphi)}{\mathrm{d}\varphi}\right]^2 > 4R(\varphi)^2\left[\frac{\mathrm{d}d(\varphi)}{\mathrm{d}\varphi}\right]^2$, then the family $C(\varphi)$ has no links.*

Proof. In this case $R(\varphi) - d(\varphi)$ is monotonic and circles are not linked because their projections are not. □

4. Construction of a fundamental domain

Theorem 4.1. *The embeddings of the triangle group defined by $< I_0, I_1, I_2(\theta) >$ are faithful and discrete for all $\pi/12 \leq \theta \leq \pi/6$.*

It follows from the following construction.

Theorem 4.2. *For each $\pi/12 \leq \theta \leq \pi/6$, there exits three surfaces $\mathbf{S}_0$, $\mathbf{S}_1$ and $\mathbf{S}_2$ such that*

— *each $\mathbf{S}_i$ is foliated by invariant $\mathbf{C}$-circles under I_i,*
— *$c_1 = \mathbf{S}_0 \cap \mathbf{S}_1$ is an invariant $\mathbf{C}$-circle under both I_0 and I_1,*
— *$c_2 = \mathbf{S}_0 \cap \mathbf{S}_2$ is an invariant $\mathbf{C}$-circle under both I_0 and $I_2(\theta)$,*
— *$\mathbf{S}_1 \cap \mathbf{S}_2$ is the point at infinity.*

Remark 4.3. *Observe that* $\mathbf{S}_0$ *is homeomorphic to an annulus. On the other hand,* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are homeomorphic to discs. The fundamental domain is homeomorphic to the solid of revolution obtained from the classical fundamental domain for the triangle modular group when it is revolved around the x-axis.*

Remark 4.4. *Those surfaces (pieces of* $\mathbf{C}$*-spheres) are called* $\mathbf{C}$*-surfaces in* [FZ, FK1]*. An application of Poincaré's theorem for complex hyperbolic geometry* [FZ, FK2] *states then that the region bounded by the three surfaces in the Heisenberg group is a fundamental domain for the triangle group. In* [FK2] *we proved the theorem for* θ *in a neighborhood of* $\pi/12$*.*

Proof. We are going to construct 3 surfaces, $\mathbf{S}_0$, $\mathbf{S}_1$, and $\mathbf{S}_2$.

— $\mathbf{S}_0$ is an annulus,
— $\mathbf{S}_1$ is a disk,
— $\mathbf{S}_2$ is a disk.

For the sake of exposition we're going to break $\mathbf{S}_1$ into a near part, $N\mathbf{S}_1$, and a far part $F\mathbf{S}_1$. The near part of $\mathbf{S}_1$ will intersect $\mathbf{S}_0$ in a circle. The far part of $\mathbf{S}_1$ is unbounded. We break up $\mathbf{S}_2$ in the same way.

To construct $\mathbf{S}_0$, we need to choose a curve in the surface of centers of $\mathbf{R}_0$. One endpoint of this curve is determined. It must be the circle c_2, which is fixed by both I_0 and I_2. Note that the other end of the curve is *not* determined because $\mathbf{R}_1$ lies in the surface of centers of $\mathbf{R}_0$. Given that one endpoint of our curve is fixed, the simplest thing to do it is to take an arc of the circle, having this endpoint, and then lift it to the surface of centers. This is the curve - call it $M(\psi)$ - which detemines $\mathbf{S}_0$. Here $\psi \in [0, \theta]$.

We would like to make $F\mathbf{S}_1$ and $F\mathbf{S}_2$ as simple as possible. So, let's make them each contained in a plane, and also make them translates of each other. So, $F\mathbf{S}_1$ and $F\mathbf{S}_2$ will be contained in parallel planes. The plane containing $F\mathbf{S}_1$ must contain $\mathbf{R}_1$ and also must be parallel to $\mathbf{R}_2$. There is a unique vertical line V which intersects both $\mathbf{R}_1$ and $\mathbf{R}_2$, and any contact plane centered on V has the desired property. Thus, we make $F\mathbf{S}_1$ by taking expanding concentric circles, about the point V intersect $\mathbf{R}_1$. All these circles lie in the desired plane. To get $F\mathbf{S}_2$ we translate $F\mathbf{S}_1$ upward until it contains points on $\mathbf{R}_2$.

So far we have not quite pinned down $F\mathbf{S}_1$ because we haven't specified the smallest circle in it. Note that $C_0(0)$ is centered on $\mathbf{R}_1$. Some dilation of $C_0(0)$ is centered at the point V intersect $\mathbf{R}_1$. We take the smallest circle in $F\mathbf{S}_1$ to be this dilation of $C_0(0)$. Once again, we take $F\mathbf{S}_2$ to be the upward translation of $F\mathbf{S}_1$.

For $N\mathbf{S}_1$ note that we already have the two ending $\mathbf{C}$-circles. One of them is $C_0(0)$ and the other one is a certain dilation of $C_0(0)$. To get the family of unlinked $\mathbf{C}$-circles interpolating these two $\mathbf{C}$-circles, we take the union of the dilations of $C_0(0)$ up to the one centered at V intersected with $\mathbf{R}_1$.

For $N\mathbf{S}_2$ note that we already have the two ending $\mathbf{C}$-circles. The small one is $C_0(\theta)$. The big one is the translation taking $V \cap \mathbf{R}1$ to $V \cap \mathbf{R}_2$ of the last dilated $C_0(0)$ centered in $V \cap \mathbf{R}_1$. That is, the big $\mathbf{C}$-circle is a translation of a dilation of $C_0(0)$. Note that the small one is a translation of a dilation of $C_0(\theta)$, where the translation and dilation are trivial. Let's examine the nontrivial translation+dilation more carefully, with a view towards interpolation between the trivial and the nontrivial. $C_0(0)$ is dilated until its center lies exactly below $\mathbf{R}_2$, and then tanslated up until it lies on $\mathbf{R}_2$. For interpolation, take any $\varphi \in [0, \theta]$. Dilate $C_0(\varphi)$ until its center lies below $\mathbf{R}_2$ and then translate until the center lies on $\mathbf{R}_2$. The union of all these circles is $N\mathbf{S}_2$.

— Technical steps —

For each fixed $\pi/12 \le \theta \le \pi/6$, the $\mathbf{R}$-circle $\mathbf{R}_2(\theta)$ is parametrized by an angle φ such that $\alpha < \varphi < \alpha + \pi$, where $\alpha = -\pi/2 + 3\theta$. We have $[\rho(\varphi)\cos(\varphi), \rho(\varphi)\sin(\varphi), Z(\varphi)]$ with

$$\rho(\varphi) = r(\theta)\sin(\alpha - \theta)/\sin(\alpha - \varphi) = \frac{\sqrt{-\cos(6\theta)}}{2\cos(3\theta - \varphi)}$$

and

$$\begin{aligned} Z(\varphi) &= \sin(2\theta) + 2r(\theta)\rho(\varphi)\sin(\theta - \varphi) \\ &= \sin(2\theta) + \frac{\cos 6\theta \sin(\theta - \varphi)}{2\cos 2\theta \cos(3\theta - \varphi)} \\ &= \frac{\sin(3\theta + \varphi)}{2\cos(3\theta - \varphi)} \end{aligned}$$

$\mathbf{R}_2(\theta)$ intersects the surface of centers of $\mathbf{R}_0$ for $\varphi = \theta$, and

$$\rho^2(\theta) = r^2(\theta) = -\frac{\cos(6\theta)}{4\cos^2(2\theta)}.$$

This intersection point determines a $\mathbf{C}$-circle c_2 which should be $\mathbf{S}_0 \cap \mathbf{S}_2$.

On the other hand any point in the x-axis is the center of an invariant circle under both I_0 and I_1. For instance, the origin is the center of the equator. To construct $\mathbf{S}_0$ we must choose one of those invariant $\mathbf{C}$-circles.

— $\mathbf{S}_0$ —

For each fixed $\pi/12 \leq \theta < \pi/6$, we choose the family of invariant $\mathbf{C}$-circles under I_0 to be the family $C_0(\psi)$, $0 \leq \psi \leq \theta$, defined by the arc in the surface of centers

$$M(\psi) = (r(\theta), \psi, \sin(2\psi)) \tag{1}$$

of constant radius. The endpoint of this arc is the invariant $\mathbf{C}$-circle c_1 under both I_0 and I_1. Its center is $(r(\theta), 0, 0)$ in the Heisenberg group. By lemma 3.1 this arc defines the surface $\mathbf{S}_0$.

— $\mathbf{S}_1$ —

The $\mathbf{C}$-circle $c_1 = C_0(0)$ will be dilated into the $\mathbf{C}$-circles $C_1(t) = l_t(c_1)$ (where l_t is Heisenberg dilation by t) whose centers belong to $\mathbf{R}_1$, and $1 \leq t \leq \frac{\rho(0)}{\rho(\theta)}$. By lemma 3.2 the dilated circles are disjoint and unlinked. Their union does not intersect $\mathbf{S}_0$, and will be part of the surface $\mathbf{S}_1$. We complete the surface $\mathbf{S}_1$ by a union of concentric $\mathbf{C}$-circles centered at the point $\rho(0)$ of the x-axis with increasing radii up to infinity: this is $F\mathbf{S}_1$. We have $c_1 = \mathbf{S}_0 \cap \mathbf{S}_1$.

— $\mathbf{S}_2$ —

Each $\mathbf{C}$-circle $C_0(\varphi)$, for $0 \leq \varphi \leq \theta$, will be dilated into a $\mathbf{C}$-circle $C_2'(\varphi) = l_{\frac{\rho(\varphi)}{\rho(\theta)}} C_0(\varphi)$ (where $l_{\frac{\rho(\varphi)}{\rho(\theta)}}$ is Heisenberg dilation by $\frac{\rho(\varphi)}{\rho(\theta)}$) whose center belongs to the curve

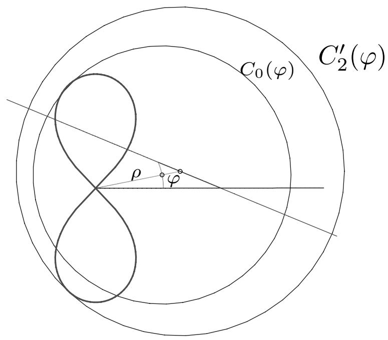
$$\mathbf{R}_2' = [\rho(\varphi)\cos(\varphi), \rho(\varphi)\sin(\varphi), \left(\frac{\rho(\varphi)}{\rho(\theta)}\right)^2 \sin(2\varphi)].$$

$\mathbf{R}_2'$ and $\mathbf{R}_2$ have same projection onto the xy-plane. But the height of $\mathbf{R}_2'$ is

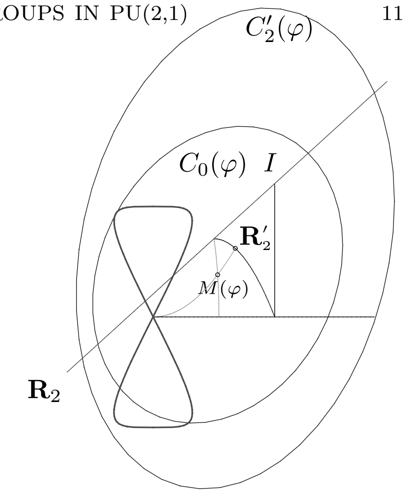
$$Z'(\varphi) = \left(\frac{\rho(\varphi)}{\rho(\theta)}\right)^2 \sin(2\varphi) = \frac{\cos^2 2\theta \sin 2\varphi}{\cos^2(3\theta - \varphi)}$$

We call ${\mathbf{S}_2}'$ the surface determined by those $\mathbf{C}$-circles.

We translate (by a vertical translation on the Heisenberg group) each $\mathbf{C}$-circle $C_2'(\varphi)$ in such a way that its center belongs to $\mathbf{R}_2$. We thus obtain a family $\mathbf{S}_2$ of invariant $\mathbf{C}$-circles $C_2(\varphi)$ for $\mathbf{R}_2$, for $0 \leq \varphi \leq \theta$. We complete the family by the union of concentric $\mathbf{C}$-circles centered at the point of $\mathbf{R}_2$ that projects over the x-axis: this is $F\mathbf{S}_2$.



Left: Projection on the z-plane of the **C**-circles of $C_0(\varphi)$ and $C_2'(\varphi)$; the second one is obtained from the first one by a dilation. In this configuration θ was fixed.



Right: A space view of the configuration. We see the curve $\mathbf{R}_2'(\varphi)$, which is a $l_{\frac{\rho(\varphi)}{\rho(\theta)}}(M(\varphi))$ where $M(\varphi)$ is defined in equation 1. Shown in the picture, the effect of the dilation at angle φ.

$\mathbf{S}_0$, $\mathbf{S}_1$ and $\mathbf{S}_2{}'$ have no linked C-circles. Any two of the **C**-circles of $\mathbf{S}_0$, $\mathbf{S}_1$ or $\mathbf{S}_2{}'$ are not linked because they are all obtained by dilation of **C**-circles of $\mathbf{S}_0$ (by lemma 3.2).

$\mathbf{S}_2$ has no linked C-circles. We will apply lemma 3.4 to the family of **C**-circles $C_2(\varphi)$. They all have centers in $\mathbf{R}_2$. Observe that $\mathbf{R}_2'$ and $\mathbf{R}_2$ project onto the same line in the xy-plane, so it is sufficient to show that the projections of the family $C_2'(\varphi)$ are disjoint. Let $\rho_0 = \rho(\theta)$, we have that $C_2'(\varphi) = l_{\frac{\rho(\varphi)}{\rho_0}} C_0(\varphi)$ has radii given by $R^2(\varphi) = \left(\frac{\rho(\varphi)}{\rho_0}\right)^2 (\rho_0^2 + \cos(2\varphi))$. Using lemma 3.4, we compute

$$\left[\frac{\mathrm{d}R^2(\varphi)}{\mathrm{d}\varphi}\right]^2 - 4R(\varphi)^2 \left[\frac{\mathrm{d}d(\varphi)}{\mathrm{d}\varphi}\right]^2$$

$$= 4\,\frac{\sin(\alpha-\theta)^4}{\sin(\varphi-\alpha)^6}\left(\cos(\varphi+\alpha)^2 + \rho_0^2\cos(2\alpha) - \rho_0^4\sin(\varphi-\alpha)^2\right)$$

But

$$\begin{aligned}\cos(\varphi+\alpha)^2+\rho_0^2\cos(2\alpha)-\rho_0^4\sin(\varphi-\alpha)^2 &\geq \cos(\varphi+\alpha)^2+{\rho_0}^2\cos(2\alpha)\\ &\qquad -\sin(\varphi-\alpha)^2\\ &= \cos(2\alpha)\left({\rho_0}^2+\cos(2\varphi)\right)\\ &\geq 0\end{aligned}$$

because $\cos(2\alpha)=\cos(6\theta-\pi)\geq 0$ for $\frac{\pi}{12}\leq\theta\leq\frac{\pi}{6}$.

$\mathbf{R}_2'$ *is below* $\mathbf{R}_2$. Let $Z'(\varphi)$ and $Z(\varphi)$ be the height of the centers of $C_2'(\varphi)$ and $C(\varphi)$. We compute

$$\begin{aligned}Z(\varphi)-Z'(\varphi) &= \frac{\sin(3\theta+\varphi)}{2\cos(3\theta-\varphi)}-\frac{\cos^2 2\theta\sin 2\varphi}{\cos^2(3\theta-\varphi)}\\ &= \frac{1}{4\cos^2(3\theta-\varphi)}\left(2\sin(3\theta+\varphi)\cos(3\theta-\varphi)-4\cos^2 2\theta\sin 2\varphi\right)\\ &= \frac{1}{4\cos^2(3\theta-\varphi)}\left(\sin 6\theta+\sin 2\varphi-4\cos^2 2\theta\sin 2\varphi\right)\\ &= \frac{1}{4\cos^2(3\theta-\varphi)}(\sin 6\theta-\frac{\sin 6\theta}{\sin 2\theta}\sin 2\varphi)\\ &= \frac{\sin 6\theta}{4\sin 2\theta\cos^2(3\theta-\varphi)}(\sin 2\theta-\sin 2\varphi)\end{aligned}$$

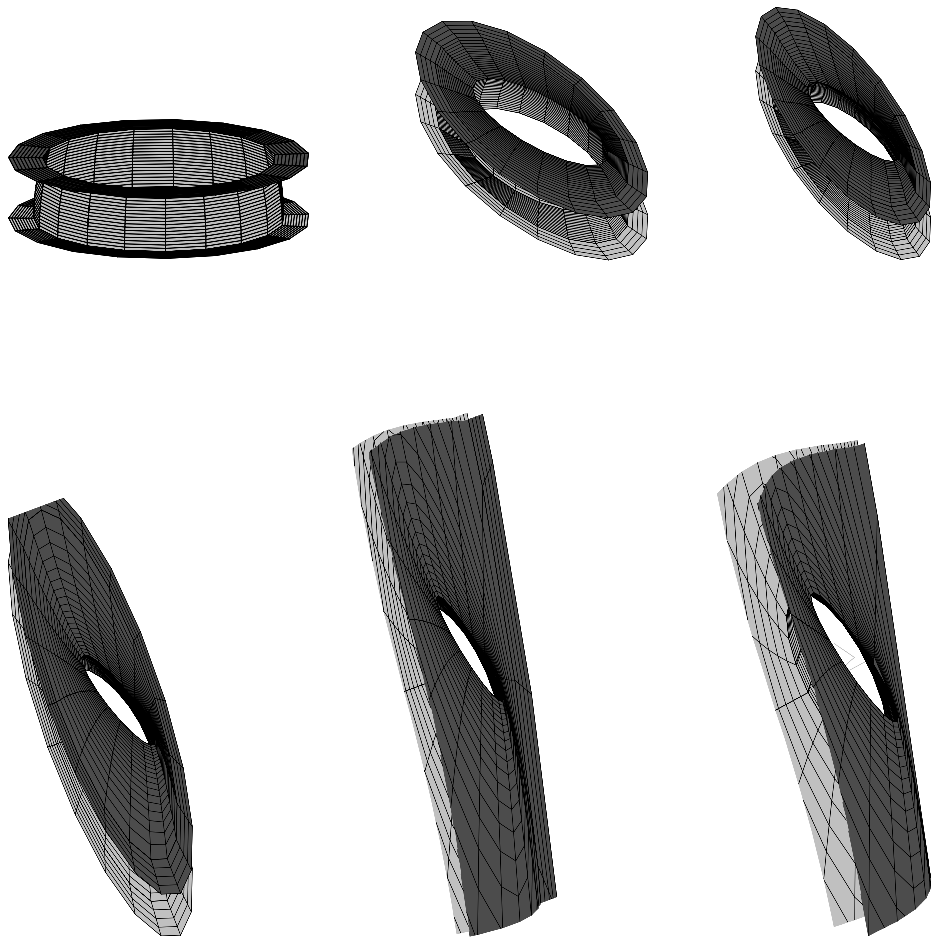
We thus deduce that $Z'(\varphi)\leq Z(\varphi), 0\leq\varphi\leq\theta$.

$\mathbf{S}_0$, $\mathbf{S}_1$ and $\mathbf{S}_2$ are disjoint. $\mathbf{S}_0$, $\mathbf{S}_1$ and ${\mathbf{S}_2}'$ are disjoint as seen previously. It remains to show that $\mathbf{S}_2$ is disjoint from the other surfaces. As $\mathbf{S}_2$ is above $S2'$, *a fortiori* $\mathbf{S}_2$ does not intersect $\mathbf{S}_0$ and $\mathbf{S}_1$.

□

Acknowledgement

We thank S. Markham and J. Parker for help with trigonometry, and R. Schwartz for improving the exposition.



Six fundamental domains for θ between $\pi/12$ and $\pi/6$. Remark that scaling is not constant. The completion of the surfaces by concentric $\mathbf{C}$-circles is partially drawn. The first upper left image corresponds to the embedding fixing a complex geodesic, while the last one fixes a totally real geodesic.

Solving the triangular Ising ferromagnet by simple mean field

GALAM, S., KOSELEFF, P.-V.
The European Physical Journal B **28** (2002), pp. 149-155

Eur. Phys. J. B **28**, 149–155 (2002)
DOI: 10.1140/epjb/e2002-00216-8

THE EUROPEAN PHYSICAL JOURNAL B

Solving the triangular Ising antiferromagnet by simple mean field

S. Galam[1,a] and P.-V. Koseleff[2]

[1] Laboratoire des Milieux Désordonnés et Hétérogènes[b], Case 86, Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05, France
[2] Équipe "Analyse Algébrique", Institut de Mathématiques[c], Case 82, Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05, France

Received 14 November 2001 / Received in final form 22 March 2002
Published online 19 July 2002 –

Abstract. Few years ago, application of the mean field Bethe scheme on a given system was shown to produce a systematic change of the system intrinsic symmetry. For instance, once applied on a ferromagnet, individual spins are no more equivalent. Accordingly a new loopwise mean field theory was designed to both go beyond the one site Weiss approach and yet preserve the initial Hamitonian symmetry. This loopwise scheme is applied here to solve the triangular antiferromagnetic Ising model. It is found to yield Wannier's exact result of no ordering at non-zero temperature. No adjustable parameter is used. Simultaneously a non-zero critical temperature is obtained for the triangular Ising ferromagnet. This simple mean field scheme opens a new way to tackle random systems.

PACS. 75.25.+z Spin arrangements in magnetically ordered materials (including neutron and spin-polarized electron studies, synchrotron-source X-ray scattering, etc.) – 05.50.+q Lattice theory and statistics (Ising, Potts, etc.) – 75.50.-y Studies of specific magnetic materials

1 Introduction

Collective phenomena are rather difficult to solve exactly. Up to date, only some one dimensional problems and the square zero field Ising model allow an exact analytical solution [1]. To compensate this situation, a rich family of approximate methods has been developed over the last hundred years. The most powerful one being the renormalization group techniques [2].

At start was the Mean Field Theory (MFT). It offers a very practical and simple tool to solve most collective phenomena [1]. While it is completely universal and generic, associated quantitative results are unusually poor. In particular critical temperatures and exponents are rather far from exact estimates [2]. Sometimes even the order of the transition may be wrong like for the instance in the Potts model [3].

The crudest and most simple version of MFT is the 1907 Weiss pioneer model [4]. It reduces the infinite number of fluctuating degrees of freedom down to one, S_i, which couples to homogeneous mean field degrees of freedom m. The thermodynamics is then solved calculating the associated partition function from which the self-consistent equation $\langle S_i \rangle = m$ (where $\langle ... \rangle$ means thermal average) is derived.

In the case of Ising systems with q nearest neighbor interactions, Weiss theory gives $\langle S_i \rangle = \tan\mathrm{h}(Kqm)$ where $K \equiv \beta J$, J is the exchange coupling, $\beta \equiv \frac{1}{k_B T}$, k_B is the Boltzmann constant and T is the temperature. Associated critical temperature is $K_c = \frac{1}{q}$. At odd with the known exact result a phase transition is obtained at $d = 1$ ($q = 2$) [1].

From there it took 28 years before Bethe improved the Weiss model [5]. Instead of just one fluctuating spin, he considers a cluster of fluctuating spins with a central one and its nearest neighbors. The main achievement of the Bethe approximation is to yield the exact result of no ordering at one dimension. However, critical temperatures given by $K_c = \tanh^{-1}(\frac{1}{q-1})$, are not much better than from Weiss model. Critical exponents stay unchanged. Latter on, using computer capabilities, larger size fluctuating clusters have been considered to obtain better critical temperatures [6].

However, a few years ago the Bethe cluster scheme was showed to systematically change the system intrinsic symmetry [7]. Starting from a system with equivalent sites like for instance a square Ising Ferromagnet, it ends up making individual sites inequivalent. At this stage it is worth to stress that an approximation can be very crude and yet not wrong as long as it preserves the intrinsic symmetry of

[a] e-mail: galam@ccr.jussieu.fr
[b] Laboratoire associé au CNRS (UMR n° 7603)
[c] Laboratoire associé au CNRS (UMR n° 7586)

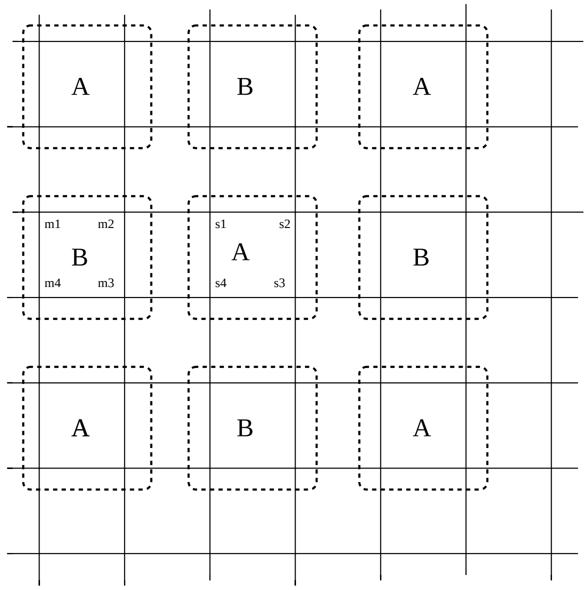


Fig. 1. The loopwise scheme in the square case: $s1$, $s2$, $s3$, $s4$ are the fluctuating spins while $m1$, $m2$, $m3$, $m4$ are mean field averages.

the problem. Otherwise it does change its physics. We are not talking here about a symmetry breaking of the higher phase symmetry as it occurs in a usual phase transition but of a change of the symmetry of the disorder phase itself.

On this basis the challenge was to find out if it is indeed possible to build a MFT which considers more than two fluctuating spins, yet preserving the initial lattice symmetry. Indeed, Galam showed it is possible using a loopwise scheme (LWS) which articulates on finite-size one-dimensional closed loops [7]. Paving the whole lattice with these loops, half of them are kept fluctuating while the other half is averaged out with mean field degrees of freedom. The scheme is illustrated in Figure 1 for the square lattice.

The LWS is a generic model. It was applied to a large class of ferromagnetic systems on Bravais lattices [7,8]. It reproduces the exact result of no ordering at one dimension. Moreover, for Ising hypercubes, it exhibits a lower critical dimension d_l for long range ordering which is equal to the Golden number $d_l = \frac{1+\sqrt{5}}{2}$. However critical exponents are unchanged from Weiss model.

On this basis, to determine the range of validity of this new LWS, it is of interest to check if it can yield new properties which are out of reach of previous mean field theories like frustration. For instance, when applied to the fully-frustrated triangular Ising antiferromagnet (TIA) most MFT predict a transition at a non-zero temperature while an earlier exact argument by Wannier proved no symmetry breaking occurs at any non-zero temperature [9].

Few years ago, to bridge this difficulty Netz and Berker introduced the hard spin recipe [10]. It combines a mean field calculation with some Monte Carlo sampling. When applied to the TIA, it yields the correct result of no ordering at $T \neq 0$. Later Banavar *et al.* suggested that the Monte Carlo sampling could be reproduced by expanding all possible products of the 6 nearest neighbors spins of the "exact spin" but it was then disproved by Netz and Berker [11].

More recently focusing on the TIA, Monroe approximated the triangular lattice with a Husimi tree built up of triangles [12]. It then allows to include properly frustration to get a correct phase diagram. However an Huzimi tree is not a triangular lattice.

In this paper we apply the very simple LWS to the fully frustrated triangular Ising antiferromagnet (TIA). The Wannier exact result is recovered [10] and a transition is found at $T = 0$. The following of the paper is organized as follows. Section 2 deals with the frustration effect. In Section 3 the LWS is presented. The TIA is solved analytically in Section 4 using the LWS. In Section 5 using the same equations, the triangular Ising ferromagnet (TIF) is also solved. Some possible applications are mentioned in the last section.

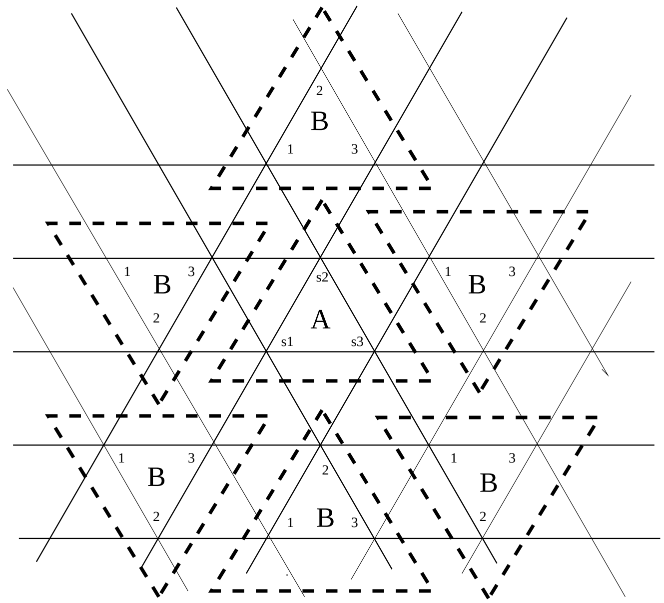


Fig. 2. The loopwise scheme in the triangular case: $s1$, $s2$, $s3$ are the fluctuating spins while 1, 2, 3 represent mean field averages $m1$, $m2$, $m3$.

2 The frustration effect

Frustration is a major ingredient of many physical systems. It results from the impossibility to minimize simultaneously all pair interactions. In turn it makes the ground state highly degenerate [9]. Frustration effects may arise from either quenched disorder or topological constraints.

Random bond spin glasses are the archetype of frustration produced by disorder. The random distribution of quenched competing interactions generates analytical difficulties in calculating the thermodynamic functions. In particular to average the disorder over the logarithm of the partition function is yet a real theoretical challenge. Usual mean field treatments failed to incorporate simultaneously frustration and quenched randomness.

On this basis the TIA has the advantage of being fully frustrated without any disorder making the study of frustration itself more easy. It is therefore the perfect candidate to check the ability of a new scheme to deal with frustration. In addition an earlier exact argument by Wannier [10] has proved the absence of symmetry breaking at any non-zero temperature for this system. At contrast most mean field like approaches produce wrongly some non-zero critical temperature. Along this line, Netz and Berker recipe [10] with Banavar *et al.* reformulation [11] stand at odd.

3 The loopwise scheme (LWS)

The LWS was introduced few years ago to overpass the symmetry inconsistency of the Bethe scheme, yet retaining its physical feature of including several fluctuating degrees of freedom [7].

To implement the LWS on any lattice requires to single out two identical interpenetrating sublattices. Each element being composed from a closed compact loop of degrees of freedom. The shape and number of these degrees of freedom are determined by the lattice topology. It is the smallest closed linear loop. For instance in the square case (Fig. 1) it includes 4 spins while for the triangular lattice (Fig. 2) 3 spins are involved. One of the sublattice is fluctuating and the other one is mean field.

Both sublattices are coupled *via* nearest neighbor interactions. The problem is thus mapped onto decoupled one-dimensional closed fluctuating chains in external fields. The fields originate from the coupling to the mean field loops. At this stage an exact analytical calculation can be performed whatever the chain size is. It is worth to note no adjustable parameter is used.

The LWS is a generic model. It was applied to a large class of ferromagnetic systems [7,8]. Being built on using closed linear loops it should be well adapted to embody frustration effects [9].

4 Solving the triangular Ising antiferromagnet

We now apply the LWS to the fully frustrated TIA. We first partition the triangular lattice into two interpenetrated triangular sublattices A and B. Thermal fluctuations are then ignored on the B-sublattice while preserved within the A-sublattice. These triangles are closed loops with no center (see Fig. 2).

All nearest neighbor (nn) plaquettes of a A-plaquette are B plaquettes, and *vice versa.* Therefore, on a given plaquette each spin has two nn spins of the same species (within the same plaquette), and four nn spins of the other species (belong respectively to three different nn plaquettes).

Above breaking of the initial lattice symmetry makes the partition function calculable by decoupling the fluctuating triangles. The A-sublattice degrees of freedom can thus be integrated out in the partition function. The initial lattice symmetry will be restored latter using the usual mean field self-consistent constraint (Eq. (4) below).

4.1 Setting the equations

Given an A plaquette, we label the 3 fluctuating spins S_1, S_2, S_3. We then introduce 3 magnetizations m_1, m_2, m_3 for corresponding B plaquettes (Fig. 2). The Hamiltonian then writes

$$H = -J(S_1S_2 + S_2S_3 + S_3S_1) - \delta J\,(S_1(m_2+m_3) + S_2(m_3+m_1) + S_3(m_1+m_2))\,, \quad (1)$$

where $\delta = 2$ accounts for the coupling to the B mean field plaquettes. From equation (1) the partition function is

$$Z = \sum_{S_i=\pm1} \exp\{-\beta H\}, \quad (2)$$

where $i = 1,\ 2,\ 3$. The three thermal averages of S_1, S_2, S_3 are given by

$$\langle S_i\rangle = \frac{1}{Z}\sum_{S_j=\pm1} S_i \exp\{-\beta H\} \ \cdot \quad (3)$$

We can thus write the associated three self-consistent equations

$$\langle S_i\rangle = m_i. \quad (4)$$

4.2 Looking for minima

Indeed we are looking for minima of the free-energy which results from the partition function Z. It is then worth to stress not all solutions of equation (4) are minima. A criterium to make equation (4) a derivative of a function is to require its cross derivatives with respect to the m_i to be equal, *i.e.*,

$$\frac{\partial}{\partial m_j}\langle S_i\rangle = \frac{\partial}{\partial m_i}\langle S_j\rangle, \quad (5)$$

for $i,\ j = 1,\ 2,\ 3$.

Writing $\mathbf{S} = (\langle S_1\rangle, \langle S_2\rangle, \langle S_3\rangle)$, and $\mathbf{m} = (m_1, m_2, m_3)$, the problem is now to find a set $S = \{\mathbf{m} \in \mathbf{R}^3;\ \mathbf{S}(\mathbf{m}) = \mathbf{m}\}$, such that there exists a function F obeying to

$$(\mathbf{m} - \mathbf{S}(\mathbf{m})) = \mathrm{d}F(\mathbf{m}) = 0\ . \quad (6)$$

To solve it, we rewrite thermal averages $\langle S_i\rangle$ as

$$\langle S_i\rangle = \frac{1}{Z}\sum_{s_i=\pm1} s_i f(s_1, s_2, s_3)\ , \quad (7)$$

where

$$f(s_1,s_2,s_3) = \exp\{K(s_1s_2 + s_1s_3 + s_2s_3) + \delta K(m_1(s_2+s_3) + m_2(s_1+s_3) + m_3(s_1+s_2))\} \ \cdot \quad (8)$$

Let $\sigma \in \Sigma_3$ be a permutation. Considering $\sigma(\mathbf{m}) = (m_{\sigma(1)}, m_{\sigma(2)}, m_{\sigma(3)})$ we have

$$\sigma(\mathbf{Z}(\mathbf{m})) = \mathbf{Z}(\sigma(\mathbf{m})), \mathbf{Z}(-\mathbf{m}) = -\mathbf{Z}(\mathbf{m})\ . \quad (9)$$

Writing $X = \exp K$ and $x_i = \exp \delta K m_i$, $\langle S_i\rangle$ are rational fractions in $(x_i,\ X)$ and we have

$$Z = \frac{D}{XT_3^2}, \quad (10)$$

$$D = (1+T_3^2)X^4 + T_2 + T_1T_3, \quad (11)$$

$$\langle S_i\rangle = 1 - 2\frac{x_i(T_1+T_3-x_i) + X^4}{D}, \quad (12)$$

where

$$T_1 = x_1 + x_2 + x_3,$$
$$T_2 = x_1x_2 + x_1x_3 + x_2x_3, \quad (13)$$
$$T_3 = x_1x_2x_3,$$

are the elementary symmetric functions. Note $D > 0$, $X > 0$, $x_i > 0$ and $|\langle S_i\rangle| < 1$.

Solving first the $K = 0$ case, we get immediately $\langle S_i\rangle = 0$ and the solution is $m_i = 0$. We can then proceed assuming $K \neq 0$.

4.3 The most general solution $m_1 \neq m_2 \neq m_3$

We can now solve the equations, starting with the most general case $m_1 \neq m_2 \neq m_3$. Equation (5) is equivalent to

$$\frac{\partial\langle S_i\rangle}{\partial x_j}\frac{\partial x_j}{\partial m_j} = \frac{\partial\langle S_j\rangle}{\partial x_i}\frac{\partial x_i}{\partial m_i}, \quad (14)$$

that is

$$\Big(x_1-x_2\Big)\Big(X^4x_3^2x_2^3x_1^3 - x_1^2x_2^2 + 2X^4x_1^2x_2^2x_3^2 + x_3^2x_2x_1 - 2X^4x_1x_2 - X^4\Big) = 0,$$
$$\Big(x_1-x_3\Big)\Big(X^4x_2^2x_3^3x_1^3 - x_3^2x_1^2 + 2X^4x_1^2x_2^2x_3^2 + x_2^2x_3x_1 - 2X^4x_3x_1 - X^4\Big) = 0,$$
$$\Big(x_2-x_3\Big)\Big(X^4x_1^2x_2^3x_3^3 - x_2^2x_3^2 + 2X^4x_1^2x_2^2x_3^2 + x_1^2x_3x_2 - 2X^4x_3x_2 - X^4\Big) = 0. \qquad (15)$$

Suppose first, three different values of m_i. It makes $x_1 \neq x_2 \neq x_3$ since $K \neq 0$, which in turn, solving equation (15) implies

$$T_1 = -\frac{T_2}{T_3}\frac{2T_3^2-1}{T_3^2-2}\ , \qquad (16)$$

and,

$$X^4 = \frac{T_2}{T_3^2-2}\cdot \qquad (17)$$

In conclusion

$$D = (1+T_3^2)X^4 + T_2 + T_1T_3, = 0 \qquad (18)$$

which is impossible since $D > 0$. Therefore, we can conclude that out of the three m_i, two must be equal. We then suppose $m_1 = m_2 \neq m_3$.

4.4 The solution exhibits the symmetry $\mathbf{m_1 = m_2 \neq m_3}$

From the above calculation we restrict the minima search to the subspace of solution $m_1 = m_2 \neq m_3$. It implies $x_1 = x_2$ and

$$\langle S_1\rangle - \langle S_3\rangle = -2\frac{x_2(1+x_1x_3)}{D}(x_1-x_3), \qquad (19)$$

so it makes

$$\frac{\exp(\delta K m_1) - \exp(\delta K m_3)}{m_1 - m_3} < 0, \qquad (20)$$

which in turn makes $K < 0$. Let us define $P \equiv x_2x_3$ and $N \equiv x_2^2x_3$, it gives

$$X^4 = \frac{P^3-N^2}{P(N^2P+2N^2-2P-1)}, \qquad (21)$$

and

$$\langle S_1\rangle + \langle S_2\rangle + \langle S_3\rangle = \frac{P-1}{P+1}\cdot \qquad (22)$$

If $P = 1$ or $N = 1$ then $m_1 = m_2 = m_3 = 0$ which is not possible since we assumed above $m_1 = m_2 \neq m_3$. So $(P-1)(N-1) \neq 0$. On the other hand, as $K < 0$, we must have

$$\frac{1-N}{m_1+m_2+m_3} > 0, \qquad (23)$$

which makes $(P-1)(N-1) < 0$ but equation (21) gives

$$X^4 = \frac{1}{P}\frac{P^3-N^2}{(1-P)+(2+P)(N^2-1)} < 0, \qquad (24)$$

which is impossible. In conclusion all the three m_i must be equal. On this basis we now assume $m_1 = m_2 = m_3 = m$.

4.5 The solution is fully symmetrical with $\mathbf{m_1 = m_2 = m_3 = m}$

We have now proved the minima belong to the solution subspace defined by the symmetry condition $m_1 = m_2 = m_3 = m$. On this basis, writing $Y = \exp(\delta K m)$, equation (4) becomes

$$m = f_K(m) = \frac{\big(Y^4-1\big)\big(\big(Y^8+Y^4+1\big)X^4+Y^4\big)}{(Y^4+1)\left((Y^8-Y^4+1)X^4+3Y^4\right)}, \qquad (25)$$

so we have either $m = 0$, or both

$$\frac{Y^4-1}{m} = \frac{e^{4\delta K m}-1}{m} > 0, \qquad (26)$$

and $K > 0$. We also deduce that $|f_K(m)| < 1$.

As $f_K(-m) = -f_K(m)$ it is enough to solve the case $m \geq 0$. We thus obtain $X > 1, Y > 1, K > 0$, or $m = 0$. Computing the derivative gives

$$f'_K(m) = 8\delta K\frac{Y^4\left(3Y^8X^8 + \left(Y^{16}+4Y^{12}+4Y^4+1\right)X^4+3Y^8\right)}{\left(Y^4+1\right)^2\left((Y^8-Y^4+1)X^4+3Y^4\right)^2}, \qquad (27)$$

$$f''_K(m) = -32\delta^2K^2\frac{Y^4(Y^4-1)}{\left(Y^4+1\right)^3\left((Y^8-Y^4+1)X^4+3Y^4\right)^3}\,g_K(m), \qquad (28)$$

where

$$g_K(m) = (Y^4+1)^6 + (2Y^8+7Y^4+2)(Y^4+1)^4(X^4-1) + (Y^{16}+7Y^{12}+21Y^8+7Y^4+1)(Y^4+1)^2 \times (X^4-1)^2 + 9Y^8(Y^8+Y^4+1)(X^4-1)^3 > 0. \qquad (29)$$

Therefore when $m \geq 0$, in addition to $0 \leq f_K(m) < 1$, we have $f'_K(m) > 0$ and $f''_K(m) < 0$. These properties allow to conclude that:

1. If $f'_K(0) \leq 1$, 0 is the only fixed point of f_K;
2. If $f'_K(0) > 1$, f_K has exactly three fixed points, $0, b, -b$ where $-1 < b < 1$.

Computing then

$$f'_K(0) = 2\delta K \frac{3\exp(4K)+1}{\exp(4K)+3}, \quad (30)$$

it appears to be an increasing function of K. It makes

$$f'_K(0) = 1, \quad (31)$$

to have a unique solution K_0. Moreover, if $K < K_0$ then $f'_K(0) < 1$ and if $K > K_0$ then $f'_K(0) > 1$.

4.6 The actual minima

Looking for minima of F_K where

$$\frac{\mathrm{d}F_K(m)}{\mathrm{d}m} = m - f_K(m) \;, \quad (32)$$

depending on the value of K, two cases appear quite naturally for $K > K_0$ and $K \leq K_0$. It shows the triangular Ising both anti and ferromagnets are solved simultaneously.

4.6.1 First case: $K > K_0$

In this case, $f_K(m) = m$ has 2 solutions $m = 0$ and $m^2 = a$ where a is a positive function of K. Having

$$F''_K(m=0) = 1 - f'_K(m=0) < 0 \;, \quad (33)$$

$m = 0$ is a maximum for F_K. In parallel

$$F''_K(m=\sqrt{a}) = F''_K(m=-\sqrt{a}) = 1 - f'_K(m_1) > 0 \;. \quad (34)$$

Therefore $m = \sqrt{a}$ and $m = -\sqrt{a}$ are minima of F_K. They correspond to the triangular Ising ferromagnet symmetry breaking at low temperatures where K_0 is the associated critical temperature.

4.6.2 Second case: $K \leq K_0$

Then the unique solution of $f_K(m) = m$ is $m = 0$. There,

$$F''_K(0) = 1 - f'_K(0) > 0, \quad (35)$$

so it is a minimum for F_K. This case embodies indeed two different physical situations.

1. The first range of positive K, $0 \leq K \leq K_0$, corresponds to the disordered phase of above triangular Ising ferromagnet.
2. At the same time, the range of negative K ($K \leq 0$) corresponds to the triangular Ising antiferromagnet. For this system the unique solution is always $m = 0$ for the whole range of temperatures $T > 0$. It means no ordering occurs for the TIA at any non zero temperature. The Wannier argument is thus recovered [10].

4.7 A transition at $T = 0$

From the exact Wannier solution the triangular Ising antiferromagnet is known to exhibit a phase transition at $T = 0$ to an ordered phase with broken symmetry among the three sublattices. Accordingly we now examine what our scheme yields in the case $K \to \infty$. To solve the equations it is more convenient to rewrite Z and $\langle S_i \rangle$ in terms of $T = \tanh(K)$ and $t_i = \tanh(\delta K m_i)$. We first note $|\langle S_i \rangle| \leq 1$ since $|\sinh(x)| \leq \cosh(x)$. Then, once the m_i are fixed within $[-1, 1]$, the condition $K \to \infty$ makes the t_i to go to either one of the three values $-1, 0, 1$.

Computing $< S_i >$ in terms of t_i and K for each one of the 27 possible limit values of the t_i set, we find 7 solutions for the m_i which are respectively

$$m_i = 0,\; i = 1, 2, 3 \quad (36)$$

and

$$m_i = m_j = -m_k = \pm 1. \quad (37)$$

To determine the actual minimum at $T = 0$ we compute the associated values for free energy $F = -k_B T \log Z$. The first solution $m_1 = m_2 = m_3 = 0$ yields

$$F = -\frac{J}{K} \log((6 + 2\exp(4K))\exp(-K)) \underset{K \to -\infty}{\longrightarrow} 1 \quad (38)$$

and for $m_1 = m_2 = 1, m_3 = -1$ we get

$$F = -\frac{J}{K} \log\left(2\exp(-K)\cos(2\delta K)(3 + \exp(4K))\right) \underset{K \to -\infty}{\longrightarrow} 1 - 2\delta, \quad (39)$$

making the solution $m_1 = m_2 = 1, m_3 = -1$ the minimum. However from equations (38, 39) the two free energies of the ordered/disordered phases are expected to become equal only at some non zero temperature, a little bit above zero temperature, that is quite close to a critical point. It is coherent to the known result of a phase transition for the triangular Ising antiferromagnet at $T = 0$ in agreement with the previous improved mean field theory by Netz and Berker [10].

5 The triangular Ising ferromagnet

Coming back to the TIF, we can go further and evaluate the value of the critical temperature K_0. At this stage it is worth to notice that all the above results are independent of the value of δ which accounts for the coupling to the mean field loops.

Since $1 \leq \dfrac{3\exp(4K)+1}{\exp(4K)+3} \leq 3$, when $K > 0$, from equation (31) we obtain

$$\frac{1}{6\delta} \leq K_0 \leq \frac{1}{2\delta} \;. \quad (40)$$

In addition, in the limit of large δ, we get

$$K_0 = \frac{1}{2\delta} - \frac{2}{\delta^2} + \frac{3}{4\delta^3} - \frac{29}{24}\frac{1}{\delta^4} + \mathcal{O}\left(\frac{1}{\delta^5}\right) \cdot \qquad (41)$$

To get a numerical estimate of the ferromagnetic critical temperature K_0 requires to have the δ value.

From equation (1) a straightforward arithmetic leads to $\delta = \frac{q-2}{2} = 2$ since 2 nn are treated exactly within the fluctuating loop out of the 6 triangular nn. Plugging then, $\delta = 2$ into equation (31) yields $K_0 = 0.1772$. It is rather far from the exact numerical estimate $K_C^e = 0.2746$ [14]. In comparison, a usual mean field gives $K_0 = \frac{1}{6} = 0.1667$, while for Bethe it is $K_0 = \tanh^{-1}(\frac{1}{5}) = 0.2027$.

6 Conclusion

In conclusion, we have showed that the very simple and generic mean field loopwise scheme, proposed by Galam [7], is able to solve exactly the triangular Ising antiferromagnet. Without any adjustable parameter it recovers the exact Wannier argument of no ordering at $T \neq 0$ and a transition at $T = 0$ [10]. From the same equations the triangular Ising ferromagnet is also solved simultaneously. A phase transition is obtained into a ferromagnetic phase at a non-zero critical temperature.

Moreover, contrary to the Bethe scheme, it preserves the initial lattice symmetry, yet going beyond the one-site Weiss approach. It also yields no transition for Ising hypercubes at $d = 1$ with a lower critical dimension of $d_l = \frac{1+\sqrt{5}}{2}$.

The loopwise scheme should allow a new solving of a very large class of physical systems, in particular random systems with frustration. For future work we consider to apply it first to the triangular Ising antiferromagnet in a finite field and then on the stacked 3D version of it. Application to the Random Field Ising model should also be done.

We would like to thank Y. Shapir and R. Netz for stimulating discussion on the manuscript.

References

1. R.K. Pathria, *Statistical Mechanics* (Pergamon Press, 1972)
2. Sh-k. Ma, *Modern Theory of Critical Phenomena* (The Benjamin Inc.: Reading MA, 1976)
3. F.Y. Wu, Rev. Mod. Phys. **54**, 235 (1982)
4. P. Weiss, J. Phys. Radium (Paris) **6**, 667 (1907)
5. H.A. Bethe, Proc. Roy. Soc. London A**150**, 552 (1935)
6. M. Suzuki, Prog. Theor. Phys. **42** 1086 (1969)
7. S. Galam, Phys. Rev. B **54**, 15991 (1996)
8. S. Galam, J. Appl. Phys. B **87**, 7040 (2000)
9. G. Toulouse, Commun. Phys. **2**, 115 (1977)
10. G.H. Wannier, Phys. Rev. **79**, 357 (1950)
11. R.R. Netz, A.N. Berker, Phys. Rev. Lett. **6**, 1377 (1991)
12. J.R. Banavar, M. Cieplak, A. Maritan, Phys. Rev. Lett. **67**, 1807 (1991) and Reply by R.R. Netz, A.N. Berker
13. J. Monroe, Physica A **256**, 217 (1998)
14. J. Adler, in *Recent developments in computer simulation studies in Condensed matter physics*, VIII, edited by D.P. Landau (Springer, 1995)